Table of Contents

Contract No.: 53-3198-0-22
MPR Reference No.: 7925-040

THE RELATIVE ACCURACY OF SAMPLE AND SHRINKAGE ESTIMATES OF STATE POVERTY RATES

FINAL REPORT

September 9, 1993

Author:

Allen L. Schirm

Submitted to:

U.S. Department of Agriculture
Food and Nutrition Service
3101 Park Center Drive
2nd Floor
Alexandria, VA 22302

Attention: Alana Landey
Project Officer

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Avenue, S.W.
Suite 550
Washington, D.C. 20024

Table of Contents

ACKNOWLEDGMENTS

I thank Bruce Klein of the Food and Nutrition Service and John Czajka of Mathematica Policy Research for helpful comments. Cara Hendricks provided exceptionally skillful assistance in conducting the simulations. Deborah Patterson assisted in the preparation of the report, and Daryl Hall edited the report.

CONTENTS

Chapter		Page
I	INTRODUCTION	1
II	STUDY DESIGN: AN OUTLINE OF THE SIMULATION PROCEDURE	9
	A. STEP 1: SPECIFY A POPULATION	9
	B. STEP 2: DRAW MULTIPLE SAMPLES FROM THE POPULATION	9
	1. Step 2a: Calculate the Sample Size for State i, $i = 1, 2, ..., 51$	10
	2. Step 2b: Draw, Without Replacement, a Simple Random Sample of Size n_i for State i, $i = 1, 2, ..., 51$	11
	3. Step 2c: Draw 1,000 Samples	11
	C. STEP 3: CALCULATE SAMPLE AND SHRINKAGE ESTIMATES	11
	1. Step 3a: Calculate the Sample Estimates	11
	2. Step 3b: Select the Best-Fitting Regression Model	12
	3. Step 3c: Calculate the Shrinkage Estimates	12
	D. STEP 4: COMPARE THE RELATIVE ACCURACY OF SAMPLE AND SHRINKAGE ESTIMATES	13
	E. POOLED SAMPLE ESTIMATION	13
III	SIMULATION RESULTS: THE RELATIVE ACCURACY OF SAMPLE AND SHRINKAGE ESTIMATES	15
	A. EVALUATING ACCURACY BY AGGREGATING ERRORS ACROSS ITERATIONS FOR EACH STATE	17
	B. EVALUATING ACCURACY BY AGGREGATING ERRORS ACROSS STATES FOR EACH ITERATION	28
	C. EVALUATING ACCURACY BY AGGREGATING ERRORS ACROSS ALL ITERATIONS AND STATES	45
	D. DISTRIBUTIONAL ACCURACY	49
	E. ACCURACY IN ESTIMATING ERROR	56
	REFERENCES	63

CONTENTS *(continued)*

Chapter		Page
	APPENDIX A - DETAILED SPECIFICATIONS FOR THE SIMULATION PROCEDURE	A-1
	APPENDIX B - ADDITIONAL TABLES OF SIMULATION RESULTS	B-1

Table of Contents

TABLES

Table		Page
III.1	POVERTY RATE ESTIMATES FROM SIMULATIONS	16
III.2	NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR IS MORE ACCURATE THAN SINGLE SAMPLE ESTIMATOR FOR A MAJORITY OF ITERATIONS	19
III.3	NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR IS MORE ACCURATE THAN POOLED SAMPLE ESTIMATOR FOR A MAJORITY OF ITERATIONS	20
III.4	NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN SINGLE SAMPLE ESTIMATOR	23
III.5	NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN POOLED SAMPLE ESTIMATOR	24
III.6	NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN SINGLE SAMPLE ESTIMATOR	26
III.7	NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN POOLED SAMPLE ESTIMATOR	27
III.8	PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR IS MORE ACCURATE THAN SINGLE SAMPLE ESTIMATOR FOR A MAJORITY OF STATES	29
III.9	PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR IS MORE ACCURATE THAN POOLED SAMPLE ESTIMATOR FOR A MAJORITY OF STATES	31
III.10	PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN SINGLE SAMPLE ESTIMATOR	33
III.11	PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN POOLED SAMPLE ESTIMATOR	34
III.12	PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN SINGLE SAMPLE ESTIMATOR	35
III.13	PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN POOLED SAMPLE ESTIMATOR	36

Table of Contents

TABLES *(continued)*

Table		Page
III.14	PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN SINGLE SAMPLE ESTIMATOR, BY WEIGHTING SCHEME USED TO CALCULATE RMSE	38
III.15	PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN POOLED SAMPLE ESTIMATOR, BY WEIGHTING SCHEME USED TO CALCULATE RMSE	39
III.16	PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN SINGLE SAMPLE ESTIMATOR, BY WEIGHTING SCHEME USED TO CALCULATE MAE	40
III.17	PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN POOLED SAMPLE ESTIMATOR, BY WEIGHTING SCHEME USED TO CALCULATE MAE	41
III.18	ACCURACY IN ESTIMATING THE NATIONAL POVERTY RATE	43
III.19	FREQUENCY DISTRIBUTION OF ABSOLUTE ERRORS IN ESTIMATING THE NATIONAL POVERTY RATE, BY ESTIMATOR	44
III.20	EFFECT OF SHRINKAGE ON ACCURACY ACROSS ALL STATES AND ITERATIONS: SHRINKAGE ESTIMATOR VERSUS SINGLE SAMPLE ESTIMATOR	46
III.21	EFFECT OF SHRINKAGE ON ACCURACY ACROSS ALL STATES AND ITERATIONS: SHRINKAGE ESTIMATOR VERSUS POOLED SAMPLE ESTIMATOR	47
III.22	FREQUENCY DISTRIBUTION OF ABSOLUTE ESTIMATION ERRORS, BY ESTIMATOR	48
III.23	ACCURACY IN ESTIMATING THE DISPERSION OF STATE ESTIMATES	51
III.24	ACCURACY IN ESTIMATING THE NUMBER OF STATES WITH POVERTY RATES BELOW 10 PERCENT OR ABOVE 13 PERCENT	53

Table of Contents

TABLES *(continued)*

Table		Page
III.25	ACCURACY IN RANKING STATES ACCORDING TO THEIR POVERTY RATES	55
III.26	ACCURACY IN IDENTIFYING THE TEN STATES WITH THE HIGHEST POVERTY RATES	57
III.27	95-PERCENT CONFIDENCE INTERVAL COVERAGE	59
III.28	DISTRIBUTION OF 95-PERCENT CONFIDENCE INTERVAL COVERAGE RATES	60

Table of Contents

EXECUTIVE SUMMARY

Policymakers have become increasingly sensitive in recent years to differences in socioeconomic conditions among regions, states, and localities. They have questioned whether the benefits of our social welfare system are shared equitably, and their concerns have intensified the need for subnational estimates for indicators of well-being and program effectiveness. Such estimates have been used to identify areas in which program participation falls shortest of need and to target resources to these areas in efforts to expand participation and improve program effectiveness. However, although accurate estimates are vital to the success of these efforts, very little is known about the relative accuracy of alternative estimators used to derive "small area" estimates of program need and effectiveness.

In this study, we assess the relative accuracy of sample and shrinkage estimates of state poverty rates. We consider both single sample and pooled sample estimators. The single sample estimator uses data for one year, and the pooled sample estimator pools data for three consecutive years. Although sample estimators are commonly used, shrinkage estimators are an attractive alternative. Shrinkage estimators calculate optimally weighted averages of estimates obtained using other methods, such as sample estimation and regression estimation. A shrinkage estimator draws on the relative strengths of the alternative estimates to obtain a better estimate. Recommended by Schirm, Swearingen, and Hendricks (1992) for estimating state poverty rates, our shrinkage estimator is an Empirical Bayes estimator that combines single sample and regression estimates. The regression estimates are obtained using a regression model that predicts state poverty rates based on such state characteristics as per capita total personal income and the proportion of the state's residents receiving Supplemental Security Income (SSI).

We use simulation methods to develop sample and shrinkage estimates of state poverty rates and to compare their relative accuracy. For our simulations, we use the March 1990 CPS sample to specify a population of individuals whose states of residence and poverty status are known. We conduct 1,000 iterations of our simulation procedure, drawing 1,000 samples from the population and calculating sample and shrinkage estimates of state poverty rates for each of the 1,000 samples. Then, we compare the estimates with the "true" state poverty rates in the population and assess the relative accuracy of the sample and shrinkage estimators.

Our principal finding is that according to a wide variety of accuracy criteria, shrinkage estimates are substantially more accurate than single or pooled sample estimates. For example, calculating root mean squared errors (RMSEs) and mean absolute errors (MAEs) for each iteration of our simulation procedure, we find that there is about a 90 to 95 percent chance that shrinkage will improve accuracy. The median reduction in the RMSE or MAE is large--about 15 to 20 percent. Shrinkage rarely decreases accuracy, and even when it does, the loss in accuracy is usually small.

In addition to evaluating the accuracy of the state poverty rate estimates, we assess the accuracy of estimated standard errors and confidence intervals as expressions of our uncertainty in the poverty rate estimates. For the pooled sample estimator, we find that the standard errors and the confidence intervals constructed from them are misleading. The standard errors are too small, and the confidence intervals are too narrow, underestimating our uncertainty and giving a false sense of accuracy. In contrast, standard errors and confidence intervals for the single sample and shrinkage estimators reflect accurately the uncertainty in estimated poverty rates.

Our simulation results provide strong evidence supporting Schirm, Swearingen, and Hendricks' (1992) recommendation to use the shrinkage estimator for estimating state poverty rates. Compared with the single and pooled sample estimators, the shrinkage estimator is almost always more accurate, and the typical gain in accuracy from shrinkage is substantial.

I. INTRODUCTION

Policymakers have become increasingly sensitive in recent years to differences in socioeconomic conditions among regions, states, and localities. They have questioned whether the benefits of our social welfare system are shared equitably, and their concerns have intensified the need for subnational estimates for indicators of well-being and program effectiveness. Such estimates can be used to identify areas in which program participation falls shortest of need and possibly to target resources to these areas in efforts to expand participation and improve program effectiveness. Such efforts have been undertaken or are under consideration for the National School Lunch Program (NSLP), the School Breakfast Program (SBP), the Child and Adult Care Food Program (CACFP), and the Special Supplemental Food Program for Women, Infants, and Children (WIC).

Although accurate estimates are vital to the success of these efforts, very little is known about the relative accuracy of alternative estimators used to derive "small area" estimates of program need and effectiveness (U.S. Office of Management and Budget 1993).[1] The leading estimators developed for small area estimation are (1) sample estimators that derive estimates directly from sample survey data, (2) model-based estimators that derive estimates using statistical models, and (3) shrinkage estimators that combine sample and model-based estimates.[2] Schirm, Swearingen, and Hendricks

[1]A small area does not have to be small or an area. The defining characteristic is a small number of sample observations--a sufficiently small number that sampling error is high. A demographic group or a large region of the country could be a small area.

[2]Estimators can also be classified as direct or indirect (U.S. Office of Management and Budget 1993). To obtain an estimate for a particular area and a particular time period, a direct estimator uses only data for that area and time period. An indirect estimator uses data from other areas or time periods. It "borrows strength" from those other areas or time periods. We examine two sample estimators in this study. One--the single sample estimator--is a direct estimator, and the other--the pooled sample estimator--is an indirect estimator. Model-based and shrinkage estimators are indirect estimators. Although shrinkage estimators are model based, they combine model estimates with sample estimates, rather than discarding the sample estimates in favor of the model estimates as do purely model-based estimators. Shrinkage estimators are sometimes called "compromise" or "composite" estimators.

(1992) examined estimators of each type and recommended a shrinkage estimator for deriving state estimates of poverty, Food Stamp Program (FSP) eligibility, and FSP participation.

In this study, we assess the relative accuracy of sample and shrinkage estimates of state poverty rates. This report documents our methods and findings. In Chapter II, we outline the simulation methods we use to develop sample and shrinkage estimates of state poverty rates and to compare their relative accuracy. In Chapter III, we present our simulation results. Our principal finding is that according to a wide variety of accuracy criteria, shrinkage estimates are substantially more accurate than sample estimates, regardless of whether the sample estimates are derived from a single sample or pooled samples. We provide more detailed specifications for our simulation procedure in Appendix A and additional tables of simulation results in Appendix B. In the remainder of this chapter, we discuss sample, model-based, and shrinkage estimators and our approach to evaluating the accuracy of alternative estimates of state poverty rates.

The basic sample estimator derives estimates directly from a single sample survey, such as one month of the Current Population Survey (CPS) or one wave of a panel of the Survey of Income and Program Participation (SIPP). Aside from its simplicity, the principal advantage of the sample estimator is that it is unbiased; that is, sample estimates are correct on average. The main disadvantage of the sample estimator is that there is often substantial sampling variability in estimates for small areas. Thus, standard errors of sample estimates are typically large.[3]

A variant of the sample estimator that has been proposed to address the high sampling error problem is the "pooled" sample estimator. Pooling combines survey data from different time periods. Plotnick (1989) and Haveman, Danziger, and Plotnick (1991) derived state poverty rate estimates by combining CPS samples for three consecutive years and dropping overlapping observations from the first and third years. This approach approximately doubles sample sizes and, therefore, reduces

[3]Recently, the Census Bureau began publishing CPS sample estimates of state poverty rates with the warning that they "should be used with caution since [they have] relatively large standard errors" (U.S. Department of Commerce 1991a).

Table of Contents

standard errors by nearly 30 percent.[4] The drawback is that a pooled estimator is biased. A state's pooled poverty rate for a single year is a weighted average of its poverty rates for three years. Because poverty rates are surely rising and falling during any three-year period, the pooled estimator is biased, although the direction of the bias cannot be determined. For estimates of year-to-year changes in poverty rates, the pooled estimator is biased downward. The pooled estimates for consecutive years incorporate two overlapping years--the second and third years pooled to obtain the first estimate are the first and second years pooled to obtain the second estimate--implying that half of the observations on which each pooled estimate is based consist of the same households whose incomes are measured at the same point in time. Because of this 50 percent overlap for which no changes in poverty status can be observed, a comparison of the two pooled poverty estimates will generally understate the year-to-year change.

Model-based estimation is an alternative to pooling as a way to reduce sampling error. The regression method is the most commonly used model-based method for small area estimation. Originally developed by Ericksen (1974), the regression method combines sample data with symptomatic information, using multivariate regression to "smooth" sample estimates, that is, to reduce their sampling variability.

The basic regression model for estimating state poverty rates is:

(1) $Y_s = XB + U,$

[4]To reduce the sampling error associated with estimates of change in monthly unemployment rates (and to reduce data collection costs), the CPS uses a "rotation group" design in which half of the selected housing units in consecutive annual samples are the same. (For monthly unemployment estimates, three-quarters of the selected housing units in consecutive monthly samples are the same.) Thus, it is necessary to pool not two but three annual CPS samples to double (approximately) the effective sample size. Half of the housing units in the middle year's sample are in the first year's sample, and the other half are in the third year's sample. The usual procedure for constructing a pooled three-year estimate--but an arbitrary choice from among several procedures--is to weight the middle year twice as heavily as each of the other two years by using all of the sample observations in the middle year and only the nonoverlapping observations in the first and third years.

Table of Contents

where Y_s is a vector of sample estimates of state poverty rates, X is a matrix containing data for each state on a set of "symptomatic indicators," and B is a vector of regression coefficients to be estimated. U is an error term reflecting both the inability of the symptomatic indicators to explain all of the interstate variation in poverty rates and the fact that the sample estimates of poverty rates are subject to sampling error. The regression estimator is:

$$(2) \quad Y_r = X\hat{B},$$

where $\hat{B}$ is the least squares estimate of B. In the regression, the state observations are often weighted by a measure of the reliability of the sample estimates.

Unlike other regression models, the regression model for deriving small area estimates has no causal interpretation. The variables on the right side of Equation (1) do not cause high or low poverty rates; instead, they are only statistically associated with high or low poverty rates and are symptomatic of differences among states. Data on symptomatic indicators are typically obtained from census or administrative records data with little or no sampling variability.

As implied by Equation (2), the regression estimates of state poverty rates are the predicted values from the regression model, where the predictions are based on the estimated regression coefficients and the observed values for the symptomatic indicators. Because of regression toward the mean, the regression estimator is biased, its principal disadvantage.[5]

Except in estimating the regression coefficients, the regression method makes no use of the sample estimates. Likewise, the sample estimator ignores the systematic relationships among state

[5] As shown by Equation (2), the regression estimates of state poverty rates lie on the estimated regression line. However, not all and maybe not any of the true poverty rates lie on that line; in other words, there is not an exact linear relationship between the poverty rates and the symptomatic indicators. Taking values only from the estimated regression line, the regression estimator smoothes away not only sampling variability, but also variability from the dispersion of the true state poverty rates about the regression line. The latter is regression toward the mean. Schirm, Swearingen, and Hendricks (1992) derive an expression for the bias of the regression estimator.

Table of Contents

poverty rates. In contrast to these estimators, shrinkage estimators seek to use all available information or, at least, the information that is most relevant and practical to use.

Shrinkage estimators calculate optimally weighted averages of estimates obtained using other methods, such as sample and regression estimates. A shrinkage estimator draws on the relative strengths of the alternative estimates to obtain a better estimate. The strength of the direct sample estimate is unbiasedness, and the strength of the model estimate is low sampling variability. A shrinkage estimator is biased by design, but such bias is accepted to reduce sampling variability. A shrinkage estimator optimally combines alternative estimates to minimize an overall measure of error, like mean squared error (MSE), that reflects both bias and sampling variability. Although a direct sample estimate may have minimum sampling error among all unbiased estimators, that minimum is typically large relative to the sampling error of some slightly biased estimator. A shrinkage estimator may offer much lower sampling error at little cost in terms of bias.

The simplest form of a shrinkage estimator is:

$$(3) \quad Y_c = aY_1 + (1 - a)Y_2,$$

where Y_c is the shrinkage (compromise) estimator that combines the alternative estimators Y_1 and Y_2, a is the vector of weights on the elements of Y_1, $(1 - a)$ is the vector of weights on the elements of Y_2, and $0 \leq a \leq 1$. To optimally combine alternative estimates, a shrinkage estimator weights the estimates according to their relative reliability. For example, a highly reliable poverty estimate is weighted more heavily and contributes more to the combined (shrinkage) poverty estimate than a less reliable poverty estimate, which is weighted less heavily and contributes less to the combined estimate. Thus, all else equal, a shrinkage estimator would place a large weight on the sample estimate for a large state and a small weight on the sample estimate for a small state.

Fay and Herriott (1979) developed a shrinkage estimator that combined sample and regression estimates of per capita income for small places (population less than 1,000) receiving funds under the

General Revenue Sharing Program. The shrinkage estimator used by Schirm, Swearingen, and Hendricks (1992) also combined sample and regression estimates. Their objective was to develop and evaluate alternative state estimates of poverty, FSP eligibility, and FSP participation. They assessed the suitability of three data sources--the census, the CPS, and SIPP--and five small area estimation methods--sample estimation, regression, the ratio correlation technique, structure preserving estimation (SPREE), and shrinkage methods. The ratio correlation technique and SPREE are model-based approaches. Based on theoretical arguments, practical considerations, and their empirical findings, Schirm, Swearingen, and Hendricks (1992) recommended deriving state estimates using CPS data and an Empirical Bayes shrinkage estimator first used for small area estimation by Ericksen and Kadane (1985, 1987).

The main limitation to Schirm, Swearingen, and Hendricks' (1992) evaluation was the lack of a fully suitable standard by which to judge the accuracy of the estimates they obtained. Although it was possible to compare estimates of sampling variability, namely, standard errors, the accuracy of the standard errors was unknown, and they shed no light on the magnitude of biases associated with the model-based and shrinkage estimators. The specification of a standard of comparison is the principal design issue for this study.

The fundamental question in evaluating accuracy is: What is the truth? Obviously, we do not know the truth; otherwise, there would be no estimation problem. Therefore, we are left with two approaches to discovering the truth. The first approach is to identify estimates that are known to be highly accurate. Although this may seem equivalent to knowing the truth, it may be that highly accurate estimates are available periodically, but much less frequently than needed. For example, although census estimates have little sampling variability even for local areas, they are available only once every 10 years. For this study, the principal difficulty with using census estimates of state poverty rates as a standard of comparison is that the nonsampling errors in census estimates are not

well-understood.[6] Because an analysis of such errors and the differences in such errors between census data and, for example, CPS data is beyond the scope of this study, we reject this first approach to ascertaining the truth. The second approach is to assume the truth and use simulation methods to derive alternative estimates under the assumed conditions. This is the approach that we have taken.

For this study, the assumed truth takes the form of a population of individuals whose states of residence and poverty status are specified. Our assumed population is based on the March 1990 CPS sample. Specifically, we take the sample as our population, ignoring the sample weights. We conduct our simulations by drawing 1,000 samples from the assumed population and calculating sample and shrinkage estimates of state poverty rates for each of the 1,000 samples. Then, we compare the estimates with the "true" state poverty rates in the assumed population and assess the relative accuracy of the sample and shrinkage estimators.

There are two potential limitations to the simulation approach. The first limitation is that the assumed truth may not accurately reflect the "real" truth. Because our assumed state poverty rates are based on a sample, they are probably more variable across states than the true poverty rates. Even if they are not more variable, our assumed poverty rates are undoubtedly different from the true poverty rates. Nevertheless, there is no reason to believe that using the true poverty rates would alter our conclusion that shrinkage estimates are more accurate than sample estimates. The second potential limitation of the simulation approach is that the sampling and estimation procedures used in the simulations may not accurately reflect the procedures that would be used in practice. As we have designed our simulations, our sampling and estimation procedures would deviate in only two ways from the procedures that would be used in practice. First, we draw simple random samples from each state's assumed population rather than attempting to mimic the complex sample design used in the CPS. Second, because of the simplified approach to sampling, we can use a well-known formula

---

[6]Eller (1992) discusses some potential sources of nonsampling error that may account for differences between census and CPS poverty and income estimates for states.

to calculate standard errors of sample estimates directly rather than estimating and using generalized variance functions. Again, although the magnitudes of the effects of shrinkage on accuracy might change, there is no reason to believe that using a complex sample design would alter our conclusion that shrinkage estimates are more accurate than sample estimates. With respect to the estimation of standard errors, Schirm, Swearingen, and Hendricks (1992) showed that even though the standard errors of the sample estimates must be used to derive the shrinkage estimates, the shrinkage estimates are not sensitive to even large errors in estimating those standard errors. We discuss these and other issues related to potential limitations of the simulation approach later in this report.

II. STUDY DESIGN: AN OUTLINE OF THE SIMULATION PROCEDURE

In this chapter, we outline our simulation procedure. This procedure has four basic steps: (1) specify a population, (2) draw multiple samples from the population, (3) calculate sample and shrinkage estimates, and (4) compare the relative accuracy of the sample and shrinkage estimates. These four steps are described in the first four sections of this chapter. In the fifth section, we describe the additions to each step required to obtain pooled sample estimates. We provide more detailed specifications for our simulation procedure in Appendix A.

A. STEP 1: SPECIFY A POPULATION

We use the March 1990 CPS sample as the population, ignoring the weights on observations and excluding unrelated individuals under age 15. This gives a total population size of approximately 158,000 individuals and state populations ranging from under 1,300 to over 14,000 across the 51 states (the 50 states and the District of Columbia). We specify the poverty status of each individual in the population using nearly the same definition employed by the Census Bureau in deriving poverty estimates from the CPS. The only difference between our definition and the Census Bureau's is minor. Instead of the poverty guidelines based on family size, number of children, and age of the family householder that are used for official government poverty estimates, we use the simplified guidelines based on family size that are used for determining eligibility for several federal programs.

B. STEP 2: DRAW MULTIPLE SAMPLES FROM THE POPULATION

In the second step of our simulation procedure, we draw multiple samples from the population specified in the first step. The purpose in drawing multiple samples is to determine how sampling variability contributes to the inaccuracy of sample and shrinkage estimates. If we drew only a single sample and discovered that the shrinkage estimates were far more accurate than the sample estimates,

we could not be sure whether the shrinkage estimator is superior or whether we had drawn an unusual sample for which the sample estimator performed unusually poorly.

Step 2 of our simulation procedure has three parts. In the first, we specify our sample design.

1. Step 2a: Calculate the Sample Size for State *i*, *i* = 1, 2, ..., 51

Replicating the complex CPS sample design in our simulations is well beyond the scope of this study. Nevertheless, we specify a sampling procedure that replicates the pattern of sampling errors found in the CPS. Specifically, we draw samples to ensure that the standard errors of the sample estimates in our simulations will generally equal or be very close to the standard errors for weighted CPS poverty rate estimates. These latter standard errors reflect the complex CPS sample design.

To simplify the simulation procedure, we use stratified simple random sampling, stratifying only by state. Given this basic sample design, the sole remaining issue is to specify the sample size for each state, that is, the number of individuals to be selected. Our expression for calculating the sample size for state i, which we derive in Appendix A, is:

$$(1) \quad n_i = \frac{T_i\,[s_i^2 + p_i\,(1 - p_i)]}{T_i\,s_i^2 + p_i\,(1 - p_i)} \,.$$

For the simulations, we set s_i equal to the standard error of the weighted CPS poverty rate estimate for state i. T_i is the population size, and p_i is the poverty rate (expressed as a proportion) in the population specified in Step 1. This p_i is the "true" poverty rate for state i in our simulations. As we show in Appendix A, the estimated standard error for a sample estimate for state i in our simulations will generally equal or be very close to s_i. Thus, the pattern of standard errors for samples estimates implied by our simple sample design is similar to the pattern of standard errors implied by the complex CPS sample design. State sample sizes in our simulations range from about 220 to over 2,200.

Table of Contents

2. Step 2b: Draw, Without Replacement, a Simple Random Sample of Size n_i for State i, $i = 1, 2, ..., 51$

The 51 state samples constitute a single national sample (henceforth, a "sample"). That sample is a stratified simple random sample. Individuals in the population are stratified by state, and independent simple random samples of individuals are drawn in each state.

3. Step 2c: Draw 1,000 Samples

We repeat Step 2b 1,000 times, drawing 1,000 independent samples. Each of the 1,000 repetitions of our simulation procedure beginning with the drawing of a sample (Step 2b) and ending with the calculation of sample and shrinkage estimates (Step 3) is an "iteration."

C. STEP 3: CALCULATE SAMPLE AND SHRINKAGE ESTIMATES

Not counting the pooled sample estimates discussed in Section E, we calculate 1,000 sets of sample and shrinkage estimates of state poverty rates, one set of 51 sample estimates and one set of 51 shrinkage estimates per iteration. To derive shrinkage estimates, we use an Empirical Bayes shrinkage estimator that combines sample and regression estimates. This estimator was used by Schirm, Swearingen, and Hendricks (1992) to derive state estimates of poverty, FSP eligibility, and FSP participation. Prior to calculating shrinkage estimates, we must calculate sample estimates and their standard errors and specify the regression model to be used.

1. Step 3a: Calculate the Sample Estimates

For state i, the sample estimate of the proportion poor is the number of individuals in the sample who are poor divided by the sample size, n_i. Expressed as a percentage, the poverty rate is the proportion poor multiplied by 100. We calculate standard errors for the sample estimates using a well-known formula for the standard error of a proportion estimated from a simple random sample drawn without replacement. This formula is displayed in Appendix A.

Table of Contents

2. Step 3b: Select the Best-Fitting Regression Model

As described in Chapter I, our regression model regresses the 51 sample estimates of state poverty rates on symptomatic indicators. The symptomatic indicators measure state characteristics that are likely to be associated with interstate differences in poverty rates. Although we do not need to calculate regression estimates prior to calculating shrinkage estimates, we do need to specify the symptomatic indicators that are included in the "best-fitting" regression model in a particular iteration. From a set of potential symptomatic indicators, we will include those for which the model obtained is parsimonious and provides a good fit. Thus, we will not include symptomatic indicators that improve the fit only marginally. We seek a model that accounts for much of the interstate variation in poverty rates with a small number of symptomatic indicators.

We allow for up to five symptomatic indicators: (1) the proportion of the state population receiving SSI, (2) state per capita total personal income, (3) the state crime rate, (4) a dummy variable equal to one for the New England states, and (5) a dummy variable equal to one if at least 1 percent of the state's total personal income is derived from the oil and gas extraction industry. Our model-fitting procedure selects the model that maximizes:

$$(2) \quad \bar{R}^2 = 1 - \left[\frac{51 - 1}{51 - k - 1}\right] (1 - R^2) ,$$

where k is the number of symptomatic indicators in the regression model (ranging from one to five), and R^2 is the usual coefficient of multiple determination. Whereas the addition of a symptomatic indicator always increases R^2, $\bar{R}^2$ will decrease if the improvement in fit, as measured by R^2, is small. We repeat our model-fitting procedure for each iteration.

3. Step 3c: Calculate the Shrinkage Estimates

We use an Empirical Bayes shrinkage estimator. This estimator was used by Ericksen and Kadane (1985, 1987) to estimate population undercounts in the 1980 census for 66 areas covering the entire U.S. and by Schirm, Swearingen, and Hendricks (1992) to estimate state poverty rates, FSP

eligibility counts, and FSP participation rates. It was originally developed by DuMouchel and Harris (1983) based on the pioneering work of Lindley and Smith (1972). The expressions for deriving the shrinkage estimates and their standard errors are given in Appendix A.

D. STEP 4: COMPARE THE RELATIVE ACCURACY OF SAMPLE AND SHRINKAGE ESTIMATES

We compare the relative accuracy of the sample and shrinkage estimates according to a wide variety of accuracy criteria, including root mean squared errors and mean absolute errors. A root mean squared error is the square root of the average squared deviation between the estimates and the true values. A mean absolute error is the average absolute deviation between the estimates and the true values. These and our other measures of accuracy are described in greater detail in Appendix A and in Chapter III. For all assessments of accuracy, the true poverty rates are the poverty rates in the population specified in Step 1.

E. POOLED SAMPLE ESTIMATION

To obtain pooled sample estimates, we must add to the first three steps of our simulation procedure. In Step 1, we must define "populations" from which to draw samples. To simulate the most often used procedure of pooling three consecutive annual samples, we use the nonoverlapping observations from the March 1989 and March 1991 CPS samples, ignoring the weights on observations and excluding unrelated individuals under age 15. From these nonoverlapping observations, we draw stratified simple random samples for each iteration. In Step 2, we draw a sample of $n_i/2$ individuals from the March 1989 CPS observations and a sample of $n_i/2$ individuals from the March 1991 CPS observations for state i. These n_i additional individuals are pooled with the n_i individuals selected from the March 1990 CPS. Thus, the pooled sample estimate is based on twice as many observations as the single sample estimate. In Step 3, the pooled sample estimate of the proportion poor is the number of individuals in the pooled sample who are poor divided by the

sample size, $2n_i$. As explained in Appendix A, we estimate the standard error for the pooled sample estimate from the standard error for the single sample estimate.

Table of Contents

III. SIMULATION RESULTS: THE RELATIVE ACCURACY OF SAMPLE AND SHRINKAGE ESTIMATES

In this chapter, we discuss our simulation results, comparing the relative accuracy of sample and shrinkage estimates of state poverty rates. Our principal finding is that shrinkage estimates are substantially more accurate than sample estimates from either a single sample or a pooled sample.

The structure of the sample and shrinkage estimates from our simulations is displayed in Table III.1. For a given state in a given iteration (represented by one cell in Table III.1), we obtain three poverty rate estimates: (1) a single sample estimate, (2) a pooled sample estimate, and (3) a shrinkage estimate. Altogether, from each of the three estimators (single sample, pooled sample, and shrinkage), we obtain 51,000 estimates--51 state estimates for each of the 1,000 iterations. Each poverty rate estimate can be compared with a true poverty rate to determine the accuracy of the estimate. For a given state, the true poverty rate remains constant across iterations and equals the poverty rate in the population specified in Step 1 of our simulation procedure, as described in Chapter II.

It is not meaningful to compare the errors in the three estimates for a single state in a single iteration. The estimates and, hence, the estimation errors may be unusual due to unusually large or small sampling errors. To control for the influence of sampling variability and discover what errors are typical, we need to aggregate estimation errors. We take three approaches to aggregating estimation errors: (1) aggregating errors across iterations for each state (summing all entries in a row in Table III.1), (2) aggregating errors across states for each iteration (summing all entries in a column in Table III.1), and (3) aggregating errors across all iterations and states (summing all entries in Table III.1).

We discuss the results obtained using these three approaches to aggregating estimation errors in Sections A, B, and C, respectively. We compare the accuracy of the single sample, pooled sample, and shrinkage estimators according to several measures of accuracy based generally on sums of errors.

Table of Contents

TABLE III.1

POVERTY RATE ESTIMATES FROM SIMULATIONS

State	Iteration 1	Iteration 2	• • •	Iteration 1,000
1. Maine	Single Sample Estimate Pooled Sample Estimate Shrinkage Estimate	Single Sample Estimate Pooled Sample Estimate Shrinkage Estimate	• • •	Single Sample Estimate Pooled Sample Estimate Shrinkage Estimate
2. New Hampshire	Single Sample Estimate Pooled Sample Estimate Shrinkage Estimate	Single Sample Estimate Pooled Sample Estimate Shrinkage Estimate	• • •	Single Sample Estimate Pooled Sample Estimate Shrinkage Estimate
• • •	• • •	• • •	• • •	• • •
51. Hawaii	Single Sample Estimate Pooled Sample Estimate Shrinkage Estimate	Single Sample Estimate Pooled Sample Estimate Shrinkage Estimate	• • •	Single Sample Estimate Pooled Sample Estimate Shrinkage Estimate

Table of Contents

In Section D, we compare how well the three estimators estimate key features of the distribution of state poverty estimates. We examine, for example, whether the dispersion in state poverty rates is represented accurately by a set of estimates and whether states are rank ordered accurately. In Section E, we compare how well the three estimators estimate error, that is, how well estimated standard errors and confidence intervals reflect the uncertainty in the poverty rate estimates.

According to the several alternative measures of accuracy we consider, we find that shrinkage estimates are substantially more accurate than single or pooled sample estimates. For example, calculating root mean squared errors (RMSEs) and mean absolute errors (MAEs) for each iteration of our simulation procedure, we find that there is about a 90 to 95 percent chance that shrinkage will improve accuracy. The median reduction in the RMSE or MAE is large--about 15 to 20 percent. Shrinkage rarely decreases accuracy, and even when it does, the loss in accuracy is usually small. Compared with the single and pooled sample estimators, the shrinkage estimator is almost always more accurate, and the typical gain in accuracy from shrinkage is substantial.

In evaluating the accuracy of estimated standard errors and confidence intervals as expressions of our uncertainty, we find that for the pooled sample estimator, the standard errors and the confidence intervals constructed from them are misleading. The standard errors are too small, and the confidence intervals are too narrow, underestimating our uncertainty and giving a false sense of accuracy. In contrast, standard errors and confidence intervals for the single sample and shrinkage estimators reflect accurately the uncertainty in estimated poverty rates.

A. EVALUATING ACCURACY BY AGGREGATING ERRORS ACROSS ITERATIONS FOR EACH STATE

One approach to measuring relative accuracy that accounts for the influence of sampling variability involves aggregating errors across iterations for each state. In other words, we can sum across the 1,000 columns for each row in Table III.1.

If we adopt this basic approach to measuring accuracy, the simplest question is: For a given state, does shrinkage improve accuracy more often than not, that is, for at least a majority of iterations? With this question in mind, our simplest measure of relative accuracy is obtained by counting the number of iterations for which the shrinkage estimate is more accurate than the sample estimate.[1] In Table III.2, we use this measure to compare the accuracy of the shrinkage and single sample estimators. Does shrinkage improve accuracy more often than not? According to Table III.2, the answer is "yes."

In Table III.2, states for which shrinkage estimates are more accurate than sample estimates for a majority of iterations are counted in the top panel, while states for which shrinkage estimates are less accurate than sample estimates for a majority of iterations are counted in the bottom panel. Thus, as labeled in Table III.2, "shrinkage increases accuracy" for the states in the top panel, and "shrinkage decreases accuracy" for the states in the bottom panel. In both panels, we display the distribution of states according to the percentage of iterations for which the shrinkage estimate is more accurate. All percentages in the top panel are above 50, while all percentages in the bottom panel are below 50.

According to Table III.2, shrinkage increases accuracy for 31 states (61 percent) and decreases accuracy for 20 states (39 percent). In the median state, the shrinkage estimate is more accurate than the sample estimate for 57 percent of the iterations.[2]

In Table III.3, we use this same measure of accuracy to compare the shrinkage and pooled sample estimators. We find that shrinkage increases accuracy for 34 states (two-thirds) and decreases

[1]The shrinkage estimate is more accurate if it is closer to the true poverty rate in absolute value.

[2]This result cannot be obtained directly from Table III.2. In Table B.1 in Appendix B, we display for each state the percentage of iterations for which the shrinkage estimate is more accurate. According to Table B.1, the maximum percentage is 96, and the minimum is 29. As we caution the reader in Appendix B, state-specific estimates are reported to show how the effects of shrinkage might vary from state to state, not to forecast the effect of shrinkage for any particular state.

Table of Contents

TABLE III.2

NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR IS MORE ACCURATE THAN SINGLE SAMPLE ESTIMATOR FOR A MAJORITY OF ITERATIONS

Effect of Shrinkage	Number of States
Shrinkage Increases Accuracy (More Accurate for a Majority of Iterations)	
Percentage of Iterations for which Shrinkage Estimate Is More Accurate than Sample Estimate:[a,b]	
> 90	3
85 - 90	7
80 - 85	3
75 - 80	2
70 - 75	1
65 - 70	2
60 - 65	3
55 - 60	7
50 - 55	3
Shrinkage Decreases Accuracy (Less Accurate for a Majority of Iterations)	
Percentage of Iterations for which Shrinkage Estimate Is More Accurate than Sample Estimate:	
45 - 50	7
40 - 45	3
35 - 40	4
30 - 35	4
≤ 30	2

[a]The shrinkage estimate is more accurate than the sample estimate if the shrinkage estimate is closer to the true poverty rate in absolute value.

[b]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

Table of Contents

TABLE III.3

NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR IS MORE ACCURATE THAN POOLED SAMPLE ESTIMATOR FOR A MAJORITY OF ITERATIONS

Effect of Shrinkage	Number of States
Shrinkage Increases Accuracy (More Accurate for a Majority of Iterations)	
Percentage of Iterations for which Shrinkage Estimate Is More Accurate than Sample Estimate:[a,b]	
> 80	5
75 - 80	4
70 - 75	2
65 - 70	5
60 - 65	5
55 - 60	7
50 - 55	6
Shrinkage Decreases Accuracy (Less Accurate for a Majority of Iterations)	
Percentage of Iterations for which Shrinkage Estimate Is More Accurate than Sample Estimate:	
45 - 50	1
40 - 45	4
35 - 40	6
30 - 35	0
≤ 30	6

[a]The shrinkage estimate is more accurate than the sample estimate if the shrinkage estimate is closer to the true poverty rate in absolute value.

[b]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

accuracy for 17 states (one-third). In the median state, the shrinkage estimate is more accurate than the pooled sample estimate for 57 percent of the iterations.[3]

A limitation to this measure of accuracy is that it ignores the relative sizes of errors except to note which is bigger. If the sample estimate is off by 2.1 percentage points, it does not matter (to this measure of accuracy) whether the shrinkage estimate is off by 0.2 or 2.0 percentage points. In both instances, the shrinkage estimate is more accurate. Such a perspective potentially understates the gains from shrinkage.

The purpose of shrinkage is to smooth out the large sampling errors. We do not expect the shrinkage estimate to be more accurate all the time and maybe not even half the time. Instead, we expect to improve accuracy by reducing substantially the occurrence of large errors while increasing the frequency of moderate errors. Moderating the largest errors, a shrinkage estimator might be preferred even if it did not increase accuracy half the time. There is no need to wrestle with this issue because our shrinkage estimator increases accuracy more often than not compared with either sample estimator.[4] Nevertheless, we will examine alternative measures of accuracy that are more informative.

We consider two measures of accuracy that fully account for the relative sizes of errors: (1) the RMSE and (2) the MAE. These measures are defined in Chapter II and Appendix A. Examining RMSEs and MAEs, we find not only improvements in accuracy for most states, but also large reductions in errors.

The RMSE, which is the square root of the mean squared error, has considerable appeal. The most widely used measures of accuracy in statistics are based on squared errors. For example, the most common method used to estimate a regression model--in any application, not just small area estimation--is least squares, which minimizes the sum of squared errors. Squaring errors penalizes

[3]According to Table B.1, the maximum percentage is 86, while the minimum is 22.

[4]We will see more explicitly in Section C that shrinkage reduces the frequency of large errors while leaving unchanged the frequency of moderate errors.

large errors heavily. If the ratio of two errors is two to one, the ratio of the squared errors is four to one.

In Table III.4, we compare the accuracy of the shrinkage and single sample estimators on the basis of RMSEs for states. States for which the shrinkage estimator has a lower RMSE are counted in the top panel of the table, while states for which the shrinkage estimator has a higher RMSE are counted in the bottom panel. Thus, as in the previous tables, shrinkage increases accuracy for the states in the top panel and decreases accuracy for the states in the bottom panel. In both panels, we display the distribution of states according to the percent change in the RMSE due to shrinkage. All the percentages in the top panel represent decreases in the RMSE, while all the percentages in the bottom panel represent increases in the RMSE. The relative accuracy of the shrinkage estimator falls as we move down in each panel.

According to Table III.4, shrinkage increases accuracy for 43 states (84 percent) and decreases accuracy for just 8 states (16 percent). In the median state, shrinkage produces a 20 percent improvement (reduction) in the RMSE.[5] For 11 states, the reduction in the RMSE exceeds 40 percent. For only 3 states does shrinkage increase the RMSE by more than 10 percent. Thus, relative to the single sample estimator, the shrinkage estimator typically increases accuracy substantially. It rarely decreases accuracy and, even then, usually only slightly.[6]

In Table III.5, we compare state RMSEs for the shrinkage and pooled sample estimators. We find that shrinkage increases accuracy for 33 states (nearly two-thirds) and decreases accuracy for 18 states (about one-third). In the median state, shrinkage produces a 14 percent improvement

---

[5]This result cannot be obtained directly from Table III.4. In Table B.2 in Appendix B, we display RMSEs for each estimator, by state. In Table B.3 in Appendix B, we display the ratio of the shrinkage estimator RMSE to the sample estimator RMSE for each state. The percentage changes in RMSEs due to shrinkage can be calculated directly from these ratios. A ratio of 0.80 indicates a 20 percent reduction in the RMSE. Although an estimator's bias is of limited relevance to an evaluation of accuracy, we display state-specific biases in Table B.4 and frequency distributions of absolute biases in Table B.5.

[6]The largest increase in the RMSE is 23 percent. The largest decrease in the RMSE is 46 percent.

TABLE III.4

NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN SINGLE SAMPLE ESTIMATOR

Effect of Shrinkage	Number of States
Shrinkage Increases Accuracy (Lowers RMSE)	
Percent Decrease in RMSE:[a]	
> 40	11
30 - 40	7
20 - 30	8
10 - 20	8
0 - 10	9
Shrinkage Decreases Accuracy (Raises RMSE)	
Percent Increase in RMSE:	
0 - 5	2
5 - 10	3
10 - 15	1
15 - 20	1
> 20	1

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

RMSE = Root Mean Squared Error

Table of Contents

TABLE III.5

NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN POOLED SAMPLE ESTIMATOR

Effect of Shrinkage	Number of States
Shrinkage Increases Accuracy (Lowers RMSE)	
Percent Decrease in RMSE:[a]	
> 50	5
40 - 50	4
30 - 40	6
20 - 30	7
10 - 20	5
0 - 10	6
Shrinkage Decreases Accuracy (Raises RMSE)	
Percent Increase in RMSE:	
0 - 10	3
10 - 20	4
20 - 30	4
30 - 40	2
40 - 50	3
> 50	2

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

RMSE = Root Mean Squared Error

Table of Contents

(reduction) in the RMSE. This average improvement over the pooled sample estimator is not as large as the average improvement over the single sample estimator. Comparing Tables III.4 and III.5 reveals that the effect of shrinkage relative to the pooled sample estimator is more variable than the effect of shrinkage relative to the single sample estimator. For many states, the shrinkage estimator is much more accurate than the pooled sample estimator, whereas for several states, the shrinkage estimator is much less accurate.[7]

An alternative to the RMSE is the MAE. In Tables III.6 and III.7, we compare the accuracy of the shrinkage estimator to the single sample and pooled sample estimators on the basis of MAEs for states. Because MAEs penalize large errors less heavily than RMSEs, the improvements in accuracy from shrinkage are generally slightly smaller when measured using MAEs instead of RMSEs. Nevertheless, shrinkage increases accuracy for well over a majority of states, and reductions in MAEs are often substantial. We find that shrinkage increases accuracy for 41 states (about 80 percent) relative to the single sample estimator and for 33 states (nearly two-thirds) relative to the pooled sample estimator.[8] In the median state, shrinkage produces a 16 percent improvement (reduction) in the MAE compared with either sample estimator.[9]

---

[7]Although the purpose of this study is to assess the relative accuracy of the sample and shrinkage estimators, we present selected results pertaining to the relative accuracy of the regression estimator. The regression estimator is described in Chapters I and II and in Appendix A. Based on state RMSEs, the regression estimator decreases accuracy for 27, 28, and 33 states compared with the single sample, pooled sample, and shrinkage estimators, respectively. In the median state, the regression estimator decreases accuracy (raises the RMSE) by 14, 10, and 27 percent relative to the single sample, pooled sample, and shrinkage estimators. There is tremendous variation about these medians, with the most extreme effects of regression reflecting large decreases in accuracy. Compared with the single sample estimator, for example, the regression estimator increases the RMSE for one state by 150 percent and decreases the RMSE for another state by over 80 percent.

[8]Although the shrinkage estimator is not much more likely to reduce accuracy when accuracy is measured by MAEs instead of RMSEs, the shrinkage estimator is more likely to reduce accuracy substantially according to MAEs.

[9]Based on state MAEs, the regression estimator decreases accuracy for 32, 29, and 33 states compared to the single sample, pooled sample, and shrinkage estimators, respectively. In the median state, the regression estimator decreases accuracy (raises the MAE) by 28, 24, and 38 percent relative to the single sample, pooled sample, and shrinkage estimators.

Table of Contents

TABLE III.6

NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN SINGLE SAMPLE ESTIMATOR

Effect of Shrinkage	Number of States
Shrinkage Increases Accuracy (Lowers MAE)	
Percent Decrease in MAE:[a]	
> 40	12
30 - 40	6
20 - 30	5
10 - 20	9
0 - 10	9
Shrinkage Decreases Accuracy (Raises MAE)	
Percent Increase in MAE:	
0 - 5	2
5 - 10	1
10 - 15	3
15 - 20	0
> 20	4

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

MAE = Mean Absolute Error

TABLE III.7

NUMBER OF STATES FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN POOLED SAMPLE ESTIMATOR

Effect of Shrinkage	Number of States
Shrinkage Increases Accuracy (Lowers MAE)	
Percent Decrease in MAE:[a]	
> 50	7
40 - 50	5
30 - 40	4
20 - 30	6
10 - 20	6
0 - 10	5
Shrinkage Decreases Accuracy (Raises MAE)	
Percent Increase in MAE:	
0 - 10	1
10 - 20	6
20 - 30	3
30 - 40	2
40 - 50	1
> 50	5

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

MAE = Mean Absolute Error

According to all three measures of accuracy that aggregate errors across iterations, the shrinkage estimator is more accurate than either sample estimator for well over a majority--anywhere from 60 to 85 percent--of the states. When we take full account of the magnitudes of errors, we find that shrinkage substantially increases accuracy. On average, state RMSEs and MAEs are reduced by about 15 to 20 percent.

B. EVALUATING ACCURACY BY AGGREGATING ERRORS ACROSS STATES FOR EACH ITERATION

In Section A, we measured accuracy by aggregating errors across iterations for each state. In this section, we measure accuracy by aggregating errors across states for each iteration. That is, we sum down the 51 rows for each column in Table III.1. We can think of different iterations as different points in time. At each point in time, we derive state estimates under conditions that have not changed except that we have drawn a different sample.

How often are shrinkage estimates more accurate? As in Section A, we consider different ways of measuring accuracy. And, as in Section A, the simplest is obtained by counting, although we now count states rather than iterations. Specifically, to determine whether the shrinkage estimator is more accurate for a particular iteration, we count the number of states for which the shrinkage estimate is more accurate than the sample estimate.[10]

With this measure of accuracy, we can answer a very simple question: For a given iteration, does shrinkage improve accuracy more often than not, that is, for at least a majority of states? In Table III.8, we use our measure that counts states to compare the accuracy of the shrinkage and single sample estimators. Does shrinkage improve accuracy more often than not? According to Table III.8, the answer is "yes."

In Table III.8, iterations for which shrinkage estimates are more accurate than sample estimates for a majority of states are counted in the top panel, while iterations for which shrinkage estimates

[10]As before, the shrinkage estimate is more accurate if it is closer to the true poverty rate in absolute value.

Table of Contents

TABLE III.8

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR IS MORE ACCURATE THAN SINGLE SAMPLE ESTIMATOR FOR A MAJORITY OF STATES

Effect of Shrinkage	Percentage of Iterations
Shrinkage Increases Accuracy (More Accurate for a Majority of States)	
Number of States for which Shrinkage Estimate Is More Accurate than Sample Estimate:[a]	
> 35	8
31 - 35	44
26 - 30	40
Shrinkage Decreases Accuracy (Less Accurate for a Majority of States)	
Number of States for which Shrinkage Estimate Is More Accurate than Sample Estimate:	
21 - 25	7
< 21	1

[a]The shrinkage estimate is more accurate than the sample estimate if the shrinkage estimate is closer to the true poverty rate in absolute value.

are less accurate than sample estimates for a majority of states are counted in the bottom panel. Thus, as labeled in Table III.8, "shrinkage increases accuracy" for the iterations in the top panel, and "shrinkage decreases accuracy" for the iterations in the bottom panel. In both panels, we display the distribution of iterations according to the number of states for which the shrinkage estimate is more accurate. The number of states is 26 or higher in the top panel and 25 or lower in the bottom panel.

According to Table III.8, shrinkage increases accuracy 92 percent of the time and decreases accuracy only 8 percent of the time. In the median iteration, the shrinkage estimate is more accurate than the sample estimate for 31 states. The most states for which shrinkage increases accuracy in any iteration is 42, and the fewest is 18.

In Table III.9, we use the same measure of accuracy to compare the shrinkage and pooled sample estimators. We find that shrinkage increases accuracy 79 percent of the time and decreases accuracy just 21 percent of the time. The median number of states with more accurate estimates from shrinkage is 28. The most states with more accurate estimates from shrinkage in any iteration is 39, and the fewest is 18.

This measure of accuracy based on counting has the same limitation when we count states as when we count iterations: it takes no account of the magnitudes of errors except to recognize that one error is bigger than another. For each iteration, we count how many states have more accurate estimates from shrinkage and how many have less accurate estimates, but we ignore how big the gains and losses in accuracy are. We would probably be willing to accept several small losses in accuracy for one or two big gains. For example, increasing the estimation error by one-tenth of a percentage point for three or four states is a good tradeoff for knocking a percentage point off a two percentage point error. We are not forced to evaluate such a tradeoff here because most of the time, shrinkage increases accuracy for more than half the states. Nevertheless, as in Section A, we will examine two alternative measures of accuracy--the RMSE and the MAE--that take into account how big the increase or decrease in accuracy is for a state.

Table of Contents

TABLE III.9

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR IS MORE ACCURATE THAN POOLED SAMPLE ESTIMATOR FOR A MAJORITY OF STATES

Effect of Shrinkage	Percentage of Iterations
Shrinkage Increases Accuracy (More Accurate for a Majority of States)	
Number of States for which Shrinkage Estimate Is More Accurate than Sample Estimate:[a]	
> 35	1
31 - 35	25
26 - 30	52
Shrinkage Decreases Accuracy (Less Accurate for a Majority of States)	
Number of States for which Shrinkage Estimate Is More Accurate than Sample Estimate:	
21 - 25	20
< 21	1

[a]The shrinkage estimate is more accurate than the sample estimate if the shrinkage estimate is closer to the true poverty rate in absolute value.

Table of Contents

In Table III.10, we compare the accuracy of the shrinkage and single sample estimators on the basis of RMSEs calculated for each iteration. According to Table III.10, shrinkage increases accuracy (reduces the RMSE) 97 percent of the time and decreases accuracy only 3 percent of the time. The median reduction in the RMSE is a very substantial 21 percent. For only 1 percent of iterations does shrinkage increase the RMSE by more than 10 percent. Thus, relative to the single sample estimator, the shrinkage estimator almost always increases accuracy substantially. It rarely decreases accuracy and almost never decreases accuracy by much.[11]

In Table III.11, we compare RMSEs for the shrinkage and pooled sample estimators. Relative to the shrinkage estimator, the pooled sample estimator performs only slightly better than the single sample estimator. According to Table III.11, shrinkage increases accuracy 90 percent of the time and decreases accuracy just 10 percent of the time. The median reduction in the RMSE is 17 percent.

In Tables III.12 and III.13, we compare the accuracy of the shrinkage estimator with the single sample and pooled sample estimators on the basis of MAEs. Although MAEs penalize large errors less heavily than RMSEs, and, thus, we might expect smaller gains in accuracy using MAEs instead of RMSEs, we find little difference from our results for RMSEs. Shrinkage almost always increases accuracy, and the gains in accuracy are typically very large.[12]

In this section, RMSEs and MAEs are calculated by adding state squared and absolute errors, respectively. This raises the issue of whether state errors should be differentially weighted. In Section A, where we calculated a RMSE for a given state by aggregating errors across iterations, the iterations were equal except for differences due entirely to sampling variability. In this section, where we calculate a RMSE for a given iteration by aggregating errors across states, states are not equal.

[11]The largest increase in the RMSE for any iteration is 19 percent. The largest decrease in the RMSE is 47 percent.

[12]Compared with the single sample estimator, shrinkage increases accuracy 97 percent of the time, and the median reduction in the MAE is 20 percent. Compared with the pooled sample estimator, shrinkage increases accuracy 90 percent of the time, and the median reduction in the MAE is 17 percent.

Table of Contents

TABLE III.10

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN SINGLE SAMPLE ESTIMATOR

Effect of Shrinkage	Percentage of Iterations
Shrinkage Increases Accuracy (Lowers RMSE)	
Percent Decrease in RMSE:[a]	
> 30	10
20 - 30	43
10 - 20	35
0 - 10	10
Shrinkage Decreases Accuracy (Raises RMSE)	
Percent Increase in RMSE:	
0 - 10	2
> 10	1

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

RMSE = Root Mean Squared Error

Table of Contents

TABLE III.11

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN POOLED SAMPLE ESTIMATOR

Effect of Shrinkage	Percentage of Iterations
Shrinkage Increases Accuracy (Lowers RMSE)	
Percent Decrease in RMSE:[a]	
> 30	8
20 - 30	28
10 - 20	36
0 - 10	18
Shrinkage Decreases Accuracy (Raises RMSE)	
Percent Increase in RMSE:	
0 - 10	8
> 10	2

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

RMSE = Root Mean Squared Error

Table of Contents

TABLE III.12

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN SINGLE SAMPLE ESTIMATOR

Effect of Shrinkage	Percentage of Iterations
Shrinkage Increases Accuracy (Lowers MAE)	
Percent Decrease in MAE:[a]	
> 30	7
20 - 30	43
10 - 20	36
0 - 10	11
Shrinkage Decreases Accuracy (Raises MAE)	
Percent Increase in MAE:	
0 - 10	2
> 10	1

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

MAE = Mean Absolute Error

TABLE III.13

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN POOLED SAMPLE ESTIMATOR

Effect of Shrinkage	Percentage of Iterations
Shrinkage Increases Accuracy (Lowers MAE)	
Percent Decrease in MAE:[a]	
> 30	8
20 - 30	30
10 - 20	34
0 - 10	19
Shrinkage Decreases Accuracy (Raises MAE)	
Percent Increase in MAE:	
0 - 10	8
> 10	2

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

MAE = Mean Absolute Error

Table of Contents

Some states are larger than others, and some have more poor people than others. Is a one percentage point error for a large state as important, more important, or less important than a one percentage point error for a small state?

We explore three weighting schemes: (1) weighting states equally, (2) weighting states by population shares, and (3) weighting states by poverty shares. A state's population share weight is the share (proportion) of all individuals in the population living in the state. A state's poverty share weight is the share (proportion) of all poor individuals in the population living in the state. The second and third weighting schemes, which are closely related, give the greatest weight to errors for states with the most people and the most poor people, respectively. These weighting schemes are described in greater detail in Appendix A.

We displayed the results obtained using the first weighting scheme, which weights states equally, in Tables III.10-III.13. In Tables III.14-III.17, we repeat those results and present the results obtained using the two differential weighting schemes. We find that the incidence of very large gains in accuracy falls wehen the errors for the large states are weighted more heavily than the errors for the small states.[13] This suggests, as expected, that the largest gains in accuracy from shrinkage are for the smallest states, where sample sizes are small and sample estimates are relatively imprecise.

The principal finding from this sensitivity analysis is that our results are not terribly sensitive to the weighting scheme used. Changing the weighting scheme barely changes either the frequency or

[13]We also find effects of weighting at the other end of the distribution. When we weighted state errors equally, we concluded that shrinkage almost never decreases accuracy by much. We draw this same conclusion from our findings pertaining to differential weighting. Moreover, although increases in RMSEs and MAEs exceeding 10 percent occur as frequently or slightly more frequently with differential weighting as with equal weighting, some of the largest decreases in accuracy are mitigated when we calculate differentially weighted RMSEs and MAEs.

TABLE III.14

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN SINGLE SAMPLE ESTIMATOR, BY WEIGHTING SCHEME USED TO CALCULATE RMSE

	Percentage of Iterations		
Effect of Shrinkage	Equal Weights	Population Weights	Poverty Weights
Shrinkage Increases Accuracy (Lowers RMSE)			
Percent Decrease in RMSE:[a]			
> 30	9	3	3
20 - 30	43	42	42
10 - 20	35	40	41
0 - 10	10	11	12
Shrinkage Decreases Accuracy (Raises RMSE)			
Percent Increase in RMSE:			
0 - 10	2	2	2
> 10	1	1	1

NOTE: A state's population weight is obtained by dividing the true state population by the true U.S. population. A state's poverty weight is obtained by dividing the true state poverty count by the true U.S. poverty count.

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

RMSE = Root Mean Squared Error

Table of Contents

TABLE III.15

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER RMSE THAN POOLED SAMPLE ESTIMATOR, BY WEIGHTING SCHEME USED TO CALCULATE RMSE

	Percentage of Iterations		
Effect of Shrinkage	Equal Weights	Population Weights	Poverty Weights
Shrinkage Increases Accuracy (Lowers RMSE)			
Percent Decrease in RMSE:[a]			
> 30	8	6	6
20 - 30	28	28	27
10 - 20	36	37	38
0 - 10	18	20	20
Shrinkage Decreases Accuracy (Raises RMSE)			
Percent Increase in RMSE:			
0 - 10	8	7	7
> 10	2	1	2

NOTE: A state's population weight is obtained by dividing the true state population by the true U.S. population. A state's poverty weight is obtained by dividing the true state poverty count by the true U.S. poverty count.

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

RMSE = Root Mean Squared Error

TABLE III.16

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN SINGLE SAMPLE ESTIMATOR, BY WEIGHTING SCHEME USED TO CALCULATE MAE

Effect of Shrinkage	Percentage of Iterations		
	Equal Weights	Population Weights	Poverty Weights
Shrinkage Increases Accuracy (Lowers MAE)			
Percent Decrease in MAE:[a]			
> 30	7	1	1
20 - 30	43	31	31
10 - 20	36	47	48
0 - 10	11	16	16
Shrinkage Decreases Accuracy (Raises MAE)			
Percent Increase in MAE:			
0 - 10	2	3	3
> 10	1	1	1

NOTE: A state's population weight is obtained by dividing the true state population by the true U.S. population. A state's poverty weight is obtained by dividing the true state poverty count by the true U.S. poverty count.

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

MAE = Mean Absolute Error

TABLE III.17

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATOR HAS LOWER MAE THAN POOLED SAMPLE ESTIMATOR, BY WEIGHTING SCHEME USED TO CALCULATE MAE

Effect of Shrinkage	Percentage of Iterations		
	Equal Weights	Population Weights	Poverty Weights
Shrinkage Increases Accuracy (Lowers MAE)			
Percent Decrease in MAE:[a]			
> 30	8	6	6
20 - 30	30	26	26
10 - 20	34	33	34
0 - 10	19	22	22
Shrinkage Decreases Accuracy (Raises MAE)			
Percent Increase in MAE:			
0 - 10	8	9	9
> 10	2	4	4

NOTE: A state's population weight is obtained by dividing the true state population by the true U.S. population. A state's poverty weight is obtained by dividing the true state poverty count by the true U.S. poverty count.

[a]The common boundary of two intervals falls in the lower interval. Thus, "10" falls in the "0 - 10" interval, not the "10 - 20" interval.

MAE = Mean Absolute Error

the average magnitude of improvement in accuracy.[14] Thus, our main conclusion is unaltered. Almost all the time, the shrinkage estimator is more accurate than the single and pooled sample estimators, and the typical gain in accuracy is substantial.[15]

In addition to adding estimation errors across states, we can add the estimates themselves across states to obtain an estimated national poverty rate for each iteration. Then, we can evaluate the accuracy of the national poverty rate estimates.

If our sole objective were to estimate the national poverty rate, we would use the single sample estimator. However, when our objective is to estimate state poverty rates, the evidence obtained so far suggests that the shrinkage estimator should be used. This raises the following question: Does shrinkage introduce substantial error into the national poverty rate estimate? The answer is "no."

In Table III.18, we compare the single sample, pooled sample, and shrinkage estimators according to several accuracy criteria, and in Table III.19, we display frequency distributions of absolute estimation errors. Although the single sample estimate is more accurate than the shrinkage estimate 62 percent of the time according to Table III.18, the errors associated with either estimator tend to be small in most iterations. According to Table III.19, roughly two-thirds to three-quarters

[14]Compared with the single sample estimator, the median reduction in the RMSE from shrinkage is 19 percent, and the median reduction in the MAE is 17 percent weighting by either the population weights or the poverty weights. The median reductions were 21 percent and 20 percent when state errors were equally weighted. Compared with the pooled sample estimator, the median reduction in the RMSE is 16 percent, and the median reduction in the MAE is 15 percent weighting by either the population weights or the poverty weights. The median reductions were both 17 percent when state errors were equally weighted. Generally, the effects of weighting are larger for the MAE than for the RMSE and for the comparison between shrinkage and single sample estimation than for the comparison between shrinkage and pooled sample estimation.

[15]We have also calculated RMSEs and MAEs for the regression estimator. Compared with the single sample estimator, the regression estimator reduces accuracy (increases the RMSE) about 95 percent of the time. The median increase in the RMSE is 21 percent. Compared with the pooled sample estimator, the regression estimator reduces accuracy in all but 5 iterations, and the median increase in the RMSE is 26 percent. Compared with the shrinkage estimator, the regression estimator always reduces accuracy. The increase in the RMSE is less than 10 percent in only 1 iteration and between 10 and 14 percent in just 14 iterations. The median increase in the RMSE is 34 percent. The regression estimator performs even more poorly according to MAEs and when state errors are weighted by population or poverty shares. Relative to the sample and shrinkage estimators, the median increase in the weighted MAE is 40 to 50 percent.

Table of Contents

TABLE III.18

ACCURACY IN ESTIMATING THE NATIONAL POVERTY RATE

Accuracy Criterion	Sample: Single	Sample: Pooled	Shrinkage
Percentage of Iterations for which Shrinkage Estimate is More Accurate than Sample Estimate[a]	38	81	n.a.
RMSE	0.184	0.383	0.212
MAE	0.146	0.360	0.171
Bias	0.005	0.360	-0.106

[a]The shrinkage estimate is more accurate than the sample estimate if the shrinkage estimate is closer to the true poverty rate in absolute value.

RMSE = Root Mean Squared Error

MAE = Mean Absolute Error

n.a. = not applicable

Table of Contents

TABLE III.19

FREQUENCY DISTRIBUTION OF ABSOLUTE ERRORS IN ESTIMATING THE NATIONAL POVERTY RATE, BY ESTIMATOR

Absolute Estimation Error (Percentage Points)[a]	Percentage of Estimates		
	Sample		
	Single	Pooled	Shrinkage
0 - 0.1	42	2	34
0.1 - 0.2	30	9	30
0.2 - 0.3	17	23	19
0.3 - 0.5	10	52	15
> 0.5	1	15	2

[a]The common boundary of two intervals falls in the lower interval. Thus, "0.1" falls in the "0 - 0.1" interval, not the "0.1 - 0.2" interval.

of the estimation errors are less than two-tenths of a percentage point, and errors rarely exceed a half percentage point. Errors for the pooled sample estimator tend to be larger; 15 percent of the time they exceed half a percentage point.[16] The MAE is nearly four-tenths of a percentage point, more than twice the MAEs for the shrinkage and single sample estimators. On balance, the shrinkage estimator is nearly as accurate as the single sample estimator for estimating the national poverty rate, and both estimators produce substantially more accurate estimates than the pooled sample estimator.

C. EVALUATING ACCURACY BY AGGREGATING ERRORS ACROSS ALL ITERATIONS AND STATES

In Section A, we measured accuracy by aggregating errors across iterations for each state. In Section B, we measured accuracy by aggregating errors across states for each iteration. In this section, we measure accuracy by aggregating errors across all iterations and states. That is, we sum across all 51,000 cells in Table III.1.

Comparing all 51,000 pairs of shrinkage and single sample estimates, we find in Table III.20 that the shrinkage estimate is more accurate 60 percent of the time. Comparing all 51,000 pairs of shrinkage and pooled sample estimates, we find in Table III.21 that the shrinkage estimate is more accurate 55 percent of the time. In Tables III.20 and III.21, we also find that shrinkage reduces RMSEs and MAEs by 15 to 20 percent compared with the single and pooled sample estimators.[17,18]

Table III.22 shows the frequency distribution of the 51,000 absolute estimation errors for each of our three estimators and helps to illustrate the effect of shrinkage. The main limitation of the

---

[16]In the 1,000 iterations, the largest absolute estimation errors are 0.64, 0.80, and 0.73 percentage points for the single sample, pooled sample, and shrinkage estimators. The median errors are 0.12, 0.36, and 0.15 percentage points.

[17]Expressions for the weighted RMSEs and MAEs are given in Appendix A.

[18]Compared with the sample estimators, the regression estimator increases RMSEs by 20 to 35 percent and MAEs by 25 to 50 percent, with the larger increases occurring when state errors are differentially weighted. Compared with the shrinkage estimator, the regression estimator increases RMSEs and MAEs by 50 to 75 percent.

Table of Contents

TABLE III.20

EFFECT OF SHRINKAGE ON ACCURACY ACROSS ALL STATES AND ITERATIONS: SHRINKAGE ESTIMATOR VERSUS SINGLE SAMPLE ESTIMATOR

Accuracy Criterion	Effect of Shrinkage
Percentage of All Shrinkage Estimates that Are More Accurate than Sample Estimates	60
Aggregate Percent Reduction in RMSE:	
• RMSE Weighted Equally	20
• RMSE Weighted by Population Shares	18
• RMSE Weighted by Poverty Shares	18
Aggregate Percent Reduction in MAE:	
• MAE Weighted Equally	19
• MAE Weighted by Population Shares	16
• MAE Weighted by Poverty Shares	16

NOTE: A state's population share weight is obtained by dividing the true state population by the true U.S. population. A state's poverty share weight is obtained by dividing the true state poverty count by the true U.S. poverty count.

RMSE = Root Mean Squared Error

MAE = Mean Absolute Error

Table of Contents

TABLE III.21

EFFECT OF SHRINKAGE ON ACCURACY ACROSS ALL STATES AND ITERATIONS: SHRINKAGE ESTIMATOR VERSUS POOLED SAMPLE ESTIMATOR

Accuracy Criterion	Effect of Shrinkage
Percentage of All Shrinkage Estimates that Are More Accurate than Sample Estimates	55
Aggregate Percent Reduction in RMSE:	
• RMSE Weighted Equally	16
• RMSE Weighted by Population Shares	16
• RMSE Weighted by Poverty Shares	16
Aggregate Percent Reduction in MAE:	
• MAE Weighted Equally	16
• MAE Weighted by Population Shares	15
• MAE Weighted by Poverty Shares	14

NOTE: A state's population share weight is obtained by dividing the true state population by the true U.S. population. A state's poverty share weight is obtained by dividing the true state poverty count by the true U.S. poverty count.

RMSE = Root Mean Squared Error

MAE = Mean Absolute Error

TABLE III.22

FREQUENCY DISTRIBUTION OF ABSOLUTE ESTIMATION ERRORS, BY ESTIMATOR

Absolute Estimation Error (Percentage Points)[a]	Percentage of All Estimates		
	Sample		
	Single	Pooled	Shrinkage
0.0 - 0.5	29	29	34
0.5 - 1.0	23	24	26
1.0 - 2.0	29	30	29
> 2.0	19	18	11

[a]The common boundary of two intervals falls in the lower interval. Thus, "1.0" falls in the "0.5 - 1.0" interval, not the "1.0 - 2.0" interval.

single sample estimator is high sampling variability. Even though the estimates are correct on average (that is, unbiased), we get very large estimation errors much too often. Using a shrinkage estimator, we are willing to accept some bias and an increase in the frequency of moderate errors to reduce the incidence of large errors. According to Table III.22, nearly 1 in 5 (single or pooled) sample estimates are more than two percentage points from the true value. In contrast, just over 1 in 10 shrinkage estimates are that far off. Such a result is not surprising. What is surprising is that while substantially decreasing the frequency of very large estimation errors, shrinkage did not increase the frequency of moderate errors. Instead, a higher proportion of errors are fairly small. While 29 percent of sample estimates are within a half percentage point of the true value, 34 percent of shrinkage estimates are that close. Furthermore, as we saw in Tables III.20 and III.21, the shrinkage estimate is more accurate than either sample estimator in pairwise comparisons 55 to 60 percent of the time.

Aggregating across both iterations and states implies the same conclusion as aggregating across either iterations or states. Shrinkage estimates are substantially more accurate than sample estimates.

D. DISTRIBUTIONAL ACCURACY

In this section, we investigate how accurately the sample and shrinkage estimates represent key features of the distribution of state poverty rates. We consider two criteria of distributional accuracy: (1) the variability of state poverty rates and (2) the rank ordering of state poverty rates.[19]

A potential limitation of the single sample estimator is that it tends to overstate variability among state poverty rates. Some states may have very low poverty rates partly because of very large negative sampling errors. Other states may have very high poverty rates partly because of very large positive

[19]For most applications, we would want to select the most accurate estimator, and our results in Sections A, B, and C are the most important for assessing accuracy. Because our findings suggest that the shrinkage estimator is substantially more accurate than the sample estimators, the main issue in this section is whether the shrinkage estimator somehow distorts the distribution of state poverty rates. That distributional accuracy is a limited standard compared with criteria such as the RMSE and MAE can be seen in a simple example. Suppose an estimator always overestimates every state's poverty rate by 10 percentage points. That terribly inaccurate estimator perfectly estimates the standard deviation and rank ordering of the state poverty rates.

sampling errors. Thus, the estimated poverty rates may be more dispersed than the true poverty rates. In contrast, a potential limitation of the shrinkage estimator is that it may understate variability among state poverty rates by shrinking the smallest and largest poverty rates toward the average poverty rate.

Because shrinkage estimates seem to be much more accurate than sample estimates according to criteria--like the RMSE and MAE--that sum state errors, we might expect that the shrinkage estimates better represent the true distribution of state poverty rates. However, if the entire gain in accuracy from shrinkage were attributable to more accurate estimates for states with moderate poverty rates, the shrinkage estimates might understate variability in poverty rates.

In Table III.23, we display results for three measures of dispersion: (1) the standard deviation, (2) the range, and (3) the interquartile range. The range is the difference between the maximum and minimum poverty rates. The interquartile range is the difference between the third and first quartiles (that is, the 75th and 25th percentiles or, roughly, the 13th and 38th highest poverty rates). In contrast to the standard deviation and, especially, the range, the interquartile range is not sensitive to one or two extreme values among the 51 state poverty rates.

In the top panel of Table III.23, we display the results pertaining to standard deviations. For each iteration, we calculated the standard deviation of the 51 single sample estimates, the standard deviation of the 51 pooled sample estimates, and the standard deviation of the 51 shrinkage estimates. Thus, we have 1,000 standard deviations of, for example, shrinkage estimates. In the last column of Table III.23, we give selected percentiles for the distribution of those 1,000 standard deviations. The 90th percentile--the value below which are 90 percent (900) of the 1,000 standard deviations--is 4.0. The 25th percentile--the value below which are 25 percent (250) of the 1,000 standard deviations--is 3.6. As shown, the true standard deviation is 3.9.

According to Table III.23, the sample estimators tend to overstate variability, while the shrinkage estimator tends to understate variability. However, the shrinkage estimator seems to more accurately

Table of Contents

TABLE III.23

ACCURACY IN ESTIMATING THE DISPERSION OF STATE ESTIMATES

	Estimated Dispersion Among State Estimates		
	Sample		
Percentile[a]	Single	Pooled	Shrinkage
Standard Deviation of State Estimates (True Value = 3.9 percent)			
90th	4.5	4.5	4.0
75th	4.4	4.4	3.9
50th (median)	4.2	4.3	3.7
25th	4.0	4.2	3.6
10th	3.9	4.0	3.4
Range of State Estimates (True Value = 20.4 percentage points)			
90th	24.1	22.4	20.7
75th	22.6	21.5	19.6
50th (median)	21.1	20.5	18.7
25th	19.7	19.5	17.7
10th	18.4	18.6	16.7
Interquartile Range of State Estimates (True Value = 4.7 percentage points)			
90th	6.0	6.0	5.3
75th	5.8	5.6	5.0
50th (median)	5.1	5.3	4.6
25th	4.7	4.9	4.2
10th	4.3	4.6	3.9

[a]For each of the 1,000 iterations, we calculated the standard deviation of, for example, the 51 state shrinkage estimates. The last column of the top panel gives the percentiles of the distribution of those 1,000 standard deviations.

reflect the dispersion in state poverty rates except when dispersion is measure by the range. We also find that the pooled sample estimator exaggerates the standard deviation and interquartile range even more than the single sample estimator. Considering the interquartile range, we find that the distribution of values for the shrinkage estimator is almost centered on the true interquartile range of 4.7. In contrast, the distributions of interquartile ranges for the single and pooled sample estimators are centered well above 4.7, with 75 percent or more of the estimated interquartile ranges above this true value.

We have also studied the dispersion in estimated poverty rates by counting the number of states with estimated poverty rates above and below specified thresholds. In our simulations, 16 states (approximately one-third) have true poverty rates below 10 percent, and 17 states (exactly one-third) have true poverty rates above 13 percent. Thus, 10 and 13 percent thresholds divide the states into approximate terciles.

In Table III.24, we determine how accurately the single sample, pooled sample, and shrinkage estimators estimate the number of states with poverty rates below our bottom threshold (10 percent) and above our top threshold (13 percent).[20] In the top panel of Table III.24, we display the distribution of the estimated number of states with poverty rates below 10 percent. According to the last column, which pertains to the shrinkage estimator, there were 16 states below the threshold in 15 percent of the iterations and 14 or 15 states below the threshold in 32 percent of the iterations.

Our findings in Table III.24 are generally consistent with our findings based on the three summary measures of dispersion. The sample estimators tend to overstate variability, while the shrinkage estimator tends to understate variability. For the sample estimators, this tendency seems to be attributable entirely to exaggerating the number of states with high poverty rates. In fact, all

---

[20]Like the results in Table III.23 for estimated ranges, the results in Table III.24 should be interpreted cautiously. The results in Table III.24 are potentially sensitive to our selection of threshold values and the patterns of estimates for one or two states. For example, our findings may have been different had we placed the lower threshold at 9.9 percent, rather than 10 percent. There would still have been 16 states with true poverty rates below the 9.9 percent threshold, but 3 of those 16 would have had poverty rates within one-tenth of a percentage point of the threshold.

Table of Contents

TABLE III.24

ACCURACY IN ESTIMATING THE NUMBER OF STATES WITH POVERTY RATES BELOW 10 PERCENT OR ABOVE 13 PERCENT

	Percentage of Iterations		
	Sample		
Number of States	Single	Pooled	Shrinkage
Number of States with Poverty Rates Below 10 Percent			
< 14	12	13	25
14 - 15	30	40	32
16	20	24	15
17 - 18	29	22	22
> 18	9	2	6
Number of States with Poverty Rates Above 13 Percent			
< 15	1	0	7
15 - 16	14	1	47
17	16	5	24
18 - 19	39	28	19
> 19	30	66	3

NOTE: Approximately one-third (16) of the states have true poverty rates below 10 percent, and exactly one-third (17) of the states have true poverty rates above 13 percent.

Table of Contents

three estimators tend to understate the number of low poverty rate states, although the shrinkage estimator is more likely to underestimate that number by at least two states. According to Table III.24, there is a strong asymmetry in errors. In estimating the number of low poverty rate states, the shrinkage estimator is off by at least two states 31 percent of the time, whereas the single and pooled sample estimators are off by at least two states 21 and 15 percent of the time. However, in estimating the number of high poverty rate states, the shrinkage estimator is off by at least two states just 10 percent of the time, whereas the single and pooled sample estimators are off by at least two states 31 and 66 percent of the time. The pooled sample estimator exaggerates the number of high poverty rate states 94 percent of the time.[21]

In assessing distributional accuracy, we have so far considered only whether estimated poverty rates are spread out too much or too little. Another relevant issue is whether the estimated poverty rates are in the right order.

For each of our three estimators, we have calculated the rank correlation between the estimated and the true state poverty rates for every iteration. The results are displayed in Table III.25. The rank correlation for the shrinkage estimator exceeds the rank correlation for the single sample estimator 88 percent of the time and the rank correlation for the pooled sample estimator 57 percent of the time. Nevertheless, all three estimators rank states fairly accurately. The minimum rank correlations exceed 0.8, and the median rank correlations exceed 0.9.

Although all three estimators rank states accurately when all 51 states are considered, a relevant question is whether the estimators rank states accurately in the tails of the poverty rate distribution.

[21]We can also aggregate across iterations. There are 18 states with true poverty rates between 10 and 13 percent. Thus, out of 51,000 estimates from a given estimator, the expected number of estimates between 10 and 13 percent is 18,000--18 in each of 1,000 iterations. The single sample estimator falls short of 18,000 by 8 percent, and the shrinkage estimator overshoots 18,000 by 8 percent. The pooled sample estimator falls short of 18,000 by 14 percent. The expected number of estimates below 10 percent is 16,000. The single sample, pooled sample, and shrinkage estimators fall short of 16,000 by 1, 4, and 6 percent. The expected number of estimates above 13 percent is 17,000. The shrinkage estimator falls short of 17,000 by 3 percent. The single and pooled sample estimators overshoot 17,000 by 9 and 19 percent.

Table of Contents

TABLE III.25

ACCURACY IN RANKING STATES ACCORDING TO THEIR POVERTY RATES

	Sample		
Accuracy Criterion	Single	Pooled	Shrinkage
Rank Correlation[a]			
Median	0.91	0.92	0.93
10th Percentile	0.87	0.90	0.90
Minimum	0.82	0.86	0.84
Percentage of Iterations for which Shrinkage Estimator Has Higher Rank Correlation than Sample Estimator	88	57	n.a.

[a]For each of the 1,000 iterations, the rank correlation between the true poverty rates and, for example, the shrinkage estimates is calculated.

n.a. = not applicable

We could imagine a federal program providing states with higher poverty rates some kind of economic assistance. However, program funding may be sufficient to assist only 10 states. How well do our estimators identify the "top 10" states--the 10 states with the highest poverty rates?

In Table III.26, we find that the shrinkage estimator is substantially more likely to identify 9 or 10 of the top 10 states than are the sample estimators. In about three-quarters of the iterations, the shrinkage estimator correctly identifies at least 9 of the 10 states with the highest poverty rates. The single and pooled sample estimators attain that standard less than half the time (in 40 and 47 percent of the iterations, respectively). Although we found earlier that the shrinkage estimator tends to underestimate the number of states with high poverty rates, that is, poverty rates above a specified threshold, it fairly accurately determines which states have high poverty rates.

E. ACCURACY IN ESTIMATING ERROR

In the previous sections of this chapter, we have assessed the relative accuracy of point estimates of state poverty rates. However, it is usual statistical practice to provide some expression of the uncertainty associated with point estimates. A conventional expression of uncertainty is an interval estimate, that is, a confidence interval.

For each of our estimators, we can calculate a confidence interval based on a point estimate and its standard error.[22] Do estimated standard errors accurately reflect the errors in our point estimates? If the standard errors do not, confidence intervals will not accurately express the range of our uncertainty.[23] In this section, we assess the accuracy of confidence intervals as expressions of our uncertainty and the error in point estimates.

---

[22]Because each of our estimators is normally distributed, the lower bound for a 95-percent confidence interval is [point estimate $-$ 1.96 $\times$ standard error], and the upper bound is [point estimate + 1.96 $\times$ standard error]. We give expressions for calculating standard errors in Appendix A.

[23]Confidence intervals may also be inaccurate, in the sense to be defined shortly, if point estimates deviate substantially from a normal distribution.

Table of Contents

TABLE III.26

ACCURACY IN IDENTIFYING THE TEN STATES WITH THE HIGHEST POVERTY RATES

Number of Top Ten States Correctly Identified	Percentage of Iterations		
	Sample		
	Single	Pooled	Shrinkage
6	2	0	0
7	14	7	1
8	44	46	24
9	36	46	58
10	4	1	16

Associated with a confidence interval is a confidence level, expressed as a percentage. A conventional confidence level is 95 percent. The frequentist interpretation of a 95-percent confidence interval is that if a 95-percent confidence interval is constructed from each of many samples using the same sampling and estimation procedures, 95 percent of the confidence intervals constructed will contain, or "cover," the true value. Hence, a 95-percent confidence interval provides 95 percent coverage.

Do 95-percent confidence intervals derived using the single sample, pooled sample, and shrinkage estimators provide 95 percent coverage? According to Table III.27, coverage is very close to 95 percent for the single sample and shrinkage estimators. For both estimators, over 93 percent of the 51,000 confidence intervals--one for each of the 51 states in each of the 1,000 iterations--contains the true poverty rate. However, for the pooled sample estimator, coverage is below 85 percent, falling substantially short of the nominal (95 percent) level.[24]

In Table III.28, we display the distribution of state coverage rates.[25] For the single sample estimator, coverage rates are very close to 95 percent, and they are above 90 percent for all 51 states. For the shrinkage estimator, coverage rates are above 90 percent for 41 states.[26] Coverage rates are between 80 and 90 percent for 6 states and between 70 and 80 percent for the other 4 states. For the pooled sample estimator, confidence interval coverage often falls far short of 95 percent. Coverage is below 60 percent for 4 states and between 60 and 70 percent for 6 states. Coverage is above 90 percent for just 27 states--barely half. The results in Tables III.27 and III.28 suggest that the standard errors for pooled estimates and the confidence intervals constructed from the standard errors are misleading. The standard errors are too small, and the confidence intervals are too narrow,

[24]For the regression estimator, coverage is only 52.8 percent. There are as many states--16--with coverage rates below 10 percent as there are states with coverage rates above 90 percent. Standard errors and confidence intervals for regression estimates are seriously misleading.

[25]In Table B.6 of Appendix B, we display the individual state coverage rates.

[26]For 16 states, the estimated confidence intervals are very conservative expressions of uncertainty, providing greater than 97 percent coverage.

Table of Contents

TABLE III.27

95-PERCENT CONFIDENCE INTERVAL COVERAGE

Coverage Criterion	Sample: Single	Sample: Pooled	Shrinkage
Percentage of All 95-Percent Confidence Intervals Including the True Value	94.4	84.3	93.2

Table of Contents

TABLE III.28

DISTRIBUTION OF 95-PERCENT CONFIDENCE INTERVAL COVERAGE RATES

Percentage of All 95-Percent Confidence Intervals Including the True Value	Number of States		
	Sample		
	Single	Pooled	Shrinkage
> 97.0	0	0	16
93.0 - 97.0	47	16	19
90.0 - 92.9	4	11	6
80.0 - 89.9	0	10	6
70.0 - 79.9	0	4	4
60.0 - 69.9	0	6	0
< 60.0	0	4	0

Table of Contents

giving a false sense of security.[27] In contrast, standard errors and confidence intervals for the single sample and shrinkage estimators generally reflect accurately the error and uncertainty in estimated poverty rates.

[27]It seems that the standard errors for pooled sample estimates are too small because when the standard errors are calculated, the observations in the pooled sample are treated as though they were obtained from a single sample. Such treatment does not take into account the bias introduced by using data from other years.

Table of Contents

REFERENCES

DuMouchel, William H., and Jeffrey E. Harris. "Bayes Methods for Combining the Results of Cancer Studies in Humans and Other Species." *Journal of the American Statistical Association*, vol. 78, no. 382, June 1983, pp. 293-315.

Eller, T.J. "State Poverty Rates and Labor Markets: 1980-1990." Paper presented at the 1992 Joint Statistical Meetings, Boston, August 1992.

Ericksen, Eugene P. "A Regression Method for Estimating Population Changes of Local Areas," *Journal of the American Statistical Assoiation,* vol. 69 no. 348, 1974, pp. 867-875.

Ericksen, Eugene P., and Joseph B. Kadane. "Sensitivity Analysis of Local Estimates of Undercount in the 1980 U.S. Census." In *Small Area Statistics: An International Symposium,* edited by R. Platek, J.N.K. Rao, C.E. Särndal, and M.P. Singh. New York: John Wiley & Sons, 1987.

Ericksen, Eugene P., and Joseph B. Kadane. "Estimating the Population in a Census Year: 1980 and Beyond." *Journal of the American Statistical Association,* vol. 80, no. 389, March 1985, pp. 98-131.

Fay, Robert E., and Roger Herriott. "Estimates of Incomes for Small-Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association,* vol. 74, no. 366, June 1979, pp. 269-277.

Haveman, Jon D., Sheldon Danziger, and Robert D. Plotnick. "State Poverty Rates for Whites, Blacks, and Hispanics in the late 1980s." *FOCUS*, vol. 13, no. 2, spring 1991, pp. 1-7.

Lindley, D.W., and A.F.M. Smith. "Bayes Estimates for the Linear Model." *Journal of the Royal Statistical Society*, ser. B, no. 34, 1972, pp. 1-41.

Plotnick, Robert D. "How Much Poverty is Reduced by State Income Transfers?" *Monthly Labor Review*, vol. 112, no. 7, July 1989, pp. 21-26.

Schirm, Allen L., Gary D. Swearingen, and Cara S. Hendricks. "Development and Evaluation of Alternative State Estimates of Poverty, Food Stamp Program Eligibility, and Food Stamp Program Participation." Washington, DC: Mathematica Policy Research, December 1992.

U.S. Department of Commerce, Bureau of the Census. "Poverty in the United States: 1990." *Current Population Reports,* series P-60, no. 175. Washington, DC: U.S. Government Printing Office, 1991a.

U.S. Department of Commerce, Bureau of the Census. *Statistical Abstract of the United States.* Washington, DC: U.S. Government Printing Office, 1991b.

U.S. Department of Commerce, Bureau of Economic Analysis. *Survey of Current Business*, vol. 71, no. 8. Washington, DC: U.S. Government Printing Office, August 1991c.

U.S. Department of Health and Human Services, Social Security Administration. *Social Security Bulletin, Annual Statistical Supplement.* Washington, DC: U.S. Government Printing Office, 1990.

U.S. Office of Management and Budget. *Indirect Estimators in Federal Programs.* Statistical Policy Working Paper no. 21. Washington, DC: OMB, July 1993.

Wolter, Kirk M. *Introduction to Variance Estimation.* New York: Springer-Verlag, 1985.

APPENDIX A

DETAILED SPECIFICATIONS FOR THE SIMULATION PROCEDURE

Table of Contents

In this appendix, we provide detailed specifications for our simulation procedure. As outlined in Chapter II, the procedure has four basic steps: (1) specify a population, (2) draw multiple samples from the population, (3) calculate sample and shrinkage estimates, and (4) compare the relative accuracy of the sample and shrinkage estimates. After discussing these four steps, we describe the additions required to obtained pooled sample estimates.

STEP 1: SPECIFY A POPULATION

We use the March 1990 CPS sample as the population, ignoring the weights on observations and excluding unrelated individuals under age 15. This gives a total population size of approximately 158,000 individuals across the 51 states (the 50 states and the District of Columbia). Except for the poverty income thresholds used, we specify the poverty status of each individual in the population using the same definition employed by the Census Bureau in deriving poverty estimates from the CPS. We compare the income of each family to a poverty threshold for that family. Individuals in each household are classified into four family types: (1) (primary) families, (2) unrelated subfamilies, (3) nonfamily householders (formerly, "primary individuals"), and (4) secondary individuals age 15 or over.[1] To determine whether a family is in poverty, we take the ratio of the family's income to the family's poverty guideline. If the ratio is less than 1.0, the family and all individuals in the family are in poverty. As noted in Chapter II, we use the simplified poverty guidelines used for determining eligibility for several federal programs as the poverty guidelines for our simulations.[2]

---

[1]A primary family and a related subfamily are treated as a single family unit, and its members fall in the first category.

[2]The guidelines depend on family size and state of residence. We averaged Office of Management and Budget (OMB) poverty income guidelines for the first and last six months of 1989 to obtain calendar year 1989 guidelines. (The annual income data collected in the March 1990 CPS pertain to 1989.) For residents of Alaska, the poverty guideline is $7,345 for a one-person family. Each additional family member increases the guideline by $2,500. For residents of Hawaii, the poverty guideline is $6,760 for a one-person family, and each additional family member increases the guideline by $2,300. For residents of the other states and the District of Columbia, the poverty guideline is $5,875 for a one-person family, and each additional family member increases the guideline by $2,000.

Table of Contents

TABLE A.1

UNWEIGHTED SAMPLE COUNTS AND POVERTY RATES FOR 1989, BY STATE

Division/State	Sample Counts: Total[a]	Sample Counts: Poor[b]	Poverty Rate[c] (Percent)
New England			
Maine	1,603	154	9.6
New Hampshire	1,340	98	7.3
Vermont	1,259	90	7.1
Massachusetts	5,745	492	8.6
Rhode Island	1,343	89	6.6
Connecticut	1,365	41	3.0
Middle Atlantic			
New York	11,687	1,625	13.9
New Jersey	6,226	492	7.9
Pennsylvania	6,488	638	9.8
East North Central			
Ohio	6,518	639	9.8
Indiana	1,769	223	12.6
Illinois	6,486	769	11.9
Michigan	6,332	795	12.6
Wisconsin	2,065	166	8.0
West North Central			
Minnesota	1,577	176	11.2
Iowa	1,884	191	10.1
Missouri	1,722	205	11.9
North Dakota	1,996	242	12.1
South Dakota	2,161	260	12.0
Nebraska	1,945	223	11.5
Kansas	1,896	190	10.0
South Atlantic			
Delaware	1,447	133	9.2
Maryland	1,532	132	8.6
District of Columbia	1,390	253	18.2
Virginia	2,326	253	10.9
West Virginia	1,841	272	14.8
North Carolina	6,105	698	11.4
South Carolina	2,175	341	15.7
Georgia	1,773	264	14.9
Florida	7,869	933	11.9

TABLE A.1 *(continued)*

Division/State	Sample Counts: Total[a]	Sample Counts: Poor[b]	Poverty Rate[c] (Percent)
East South Central			
Kentucky	1,630	258	15.8
Tennessee	1,812	300	16.6
Alabama	1,860	320	17.2
Mississippi	2,063	455	22.1
West South Central			
Arkansas	2,000	336	16.8
Louisiana	1,525	357	23.4
Oklahoma	1,774	233	13.1
Texas	8,772	1,646	18.8
Mountain			
Montana	2,035	293	14.4
Idaho	2,093	245	11.7
Wyoming	1,417	138	9.7
Colorado	1,690	198	11.7
New Mexico	2,459	432	17.6
Arizona	1,886	268	14.2
Utah	1,949	142	7.3
Nevada	1,614	158	9.8
Pacific			
Washington	1,835	173	9.4
Oregon	1,609	178	11.1
California	14,413	1,918	13.3
Alaska	2,122	240	11.3
Hawaii	1,515	193	12.7

SOURCE: March 1990 Current Population Survey.

[a]The state totals are the "true" state population sizes in the simulations.

[b]The counts of poor persons are the "true" state poverty counts in the simulations.

[c]The unweighted poverty rates are the "true" poverty rates in the simulations.

Table of Contents

TABLE A.2

WEIGHTED AND UNWEIGHTED POVERTY RATES FOR 1989, BY STATE

Division/State	1989 Poverty Rate (Percent) Weighted	Unweighted[a]
New England		
Maine	9.5	9.6
New Hampshire	7.2	7.3
Vermont	7.1	7.1
Massachusetts	8.1	8.6
Rhode Island	6.4	6.6
Connecticut	2.9	3.0
Middle Atlantic		
New York	12.3	13.9
New Jersey	7.5	7.9
Pennsylvania	9.8	9.8
East North Central		
Ohio	9.9	9.8
Indiana	13.2	12.6
Illinois	12.0	11.9
Michigan	12.6	12.6
Wisconsin	8.0	8.0
West North Central		
Minnesota	11.2	11.2
Iowa	9.9	10.1
Missouri	11.2	11.9
North Dakota	11.8	12.1
South Dakota	12.6	12.0
Nebraska	11.6	11.5
Kansas	10.3	10.0
South Atlantic		
Delaware	9.2	9.2
Maryland	8.6	8.6
District of Columbia	18.0	18.2
Virginia	10.8	10.9
West Virginia	14.9	14.8
North Carolina	11.6	11.4
South Carolina	16.3	15.7
Georgia	14.2	14.9
Florida	11.8	11.9

Table of Contents

found in the CPS. Specifically, if s_i is the standard error--calculated to reflect the complex CPS sample design--for the weighted CPS poverty rate estimate for state i, we draw samples to ensure that the standard errors of the sample estimates in our simulations will generally equal or be very close to s_i. Thus, while simplifying our simulation procedures, we can mimic the outcome of the procedures that are used in the CPS and make our simulations realistic.

To simplify the simulation procedure, we use stratified simple random sampling and stratify only by state. Within strata, we sample without replacement. Given this basic sample design, we need to specify only the sample size for each state, that is, the number of individuals to be selected. Our expression for calculating the sample size for state i, displayed in Chapter II, can be derived easily.

Under the sample design specified for our simulations, we draw, without replacement, a simple random sample for each state. Suppose we have obtained a sample estimate of the poverty rate for state i. An unbiased estimator of the standard error for that poverty rate is:

$$(1) \quad \hat{s}_i = \sqrt{\left(1 - \frac{n_i}{T_i}\right) \frac{\hat{p}_i\,(1 - \hat{p}_i)}{(n_i - 1)}}\ ,$$

where n_i is the sample size for state i, T_i is the population size, and $\hat{p}_i$ is the estimated poverty rate (expressed as a proportion). Squaring both sides of this expression and solving for the sample size gives:

$$(2) \quad n_i = \frac{T_i\,[\hat{s}_i^2 + \hat{p}_i\,(1 - \hat{p}_i)]}{T_i\,\hat{s}_i^2 + \hat{p}_i\,(1 - \hat{p}_i)}.$$

For the simulations, we set $\hat{s}_i$ equal to s_i, the standard error of the weighted CPS poverty rate estimate for state i. We set $\hat{p}_i$ equal to p_i, the poverty rate (expressed as a proportion) in the population specified in Step 1. This p_i is the "true" poverty rate for state i in our simulations. Thus, as given in Chapter II, our expression for calculating the sample size for state i is:

Table of Contents

TABLE A.3

CALCULATING STATE SAMPLE SIZES FOR SIMULATIONS

Division/State	Assumed Population Size	Assumed Poverty Rate (Percent)	Target Standard Error[a] (Percent)	Sample Size for Simulations[b]
New England				
Maine	1,603	9.6	1.6	285
New Hampshire	1,340	7.3	1.6	232
Vermont	1,259	7.1	1.5	234
Massachusetts	5,745	8.6	0.8	1,052
Rhode Island	1,343	6.6	1.5	239
Connecticut	1,365	3.0	1.0	233
Middle Atlantic				
New York	11,687	13.9	0.7	2,101
New Jersey	6,226	7.9	0.7	1,105
Pennsylvania	6,488	9.8	0.8	1,141
East North Central				
Ohio	6,518	9.8	0.8	1,099
Indiana	1,769	12.6	1.9	270
Illinois	6,486	11.9	0.9	1,056
Michigan	6,332	12.6	0.9	1,082
Wisconsin	2,065	8.0	1.4	331
West North Central				
Minnesota	1,577	11.2	1.7	277
Iowa	1,884	10.1	1.5	325
Missouri	1,722	11.9	1.7	297
North Dakota	1,996	12.1	1.6	349
South Dakota	2,161	12.0	1.6	359
Nebraska	1,945	11.5	1.6	331
Kansas	1,896	10.0	1.6	310
South Atlantic				
Delaware	1,447	9.2	1.7	250
Maryland	1,532	8.6	1.6	257
District of Columbia	1,390	18.2	2.4	217
Virginia	2,326	10.9	1.5	383
West Virginia	1,841	14.8	1.9	297
North Carolina	6,105	11.4	0.9	1,074
South Carolina	2,175	15.7	1.8	355
Georgia	1,773	14.9	1.8	308
Florida	7,869	11.9	0.8	1,235

TABLE A.2 *(continued)*

Division/State	1989 Poverty Rate (Percent) Weighted	Unweighted[a]
East South Central		
Kentucky	15.7	15.8
Tennessee	16.7	16.6
Alabama	17.9	17.2
Mississippi	21.7	22.1
West South Central		
Arkansas	17.7	16.8
Louisiana	23.1	23.4
Oklahoma	12.9	13.1
Texas	16.0	18.8
Mountain		
Montana	14.5	14.4
Idaho	11.7	11.7
Wyoming	9.3	9.7
Colorado	10.8	11.7
New Mexico	17.1	17.6
Arizona	12.7	14.2
Utah	6.9	7.3
Nevada	9.6	9.8
Pacific		
Washington	9.2	9.4
Oregon	10.8	11.1
California	12.0	13.3
Alaska	13.0	11.3
Hawaii	12.2	12.7

SOURCE: March 1990 Current Population Survey.

[a]The unweighted poverty rates are the "true" poverty rates in the simulations.

In Table A.1, we display the unweighted state sample counts and poverty rates obtained from the March 1990 CPS. The sample counts in the "Total" column give the state population sizes in our simulations. The unweighted poverty rates are the "true" poverty rates in our simulations. In Table A.2, we display weighted and unweighted state poverty rates for 1989 estimated from the March 1990 CPS. Although there are differences--generally small--between the weighted and unweighted poverty rates for individual states, the two sets of rates are similarly centered and dispersed, and their rank correlation is 0.98.[3] Hence, the distribution of unweighted poverty rates, which serve as the true rates in our simulations, is very similar to the distribution of weighted poverty rates.[4]

STEP 2: DRAW MULTIPLE SAMPLES FROM THE POPULATION

In the second step of our simulation procedure, we draw multiple samples from the population specified in the first step. The purpose in drawing multiple samples is to determine how sampling variability contributes to the inaccuracy of sample and shrinkage estimates. If we drew only a single sample and discovered that the shrinkage estimates were far more accurate than the sample estimates, we could not be sure whether the shrinkage estimator is generally more accurate or whether we had drawn an unusual sample for which the sample estimator performed unusually poorly. Step 2 of our simulation procedure has three parts.

Step 2a: Calculate the Sample Size for State i, $i = 1, 2, ..., 51$

Replicating the complex CPS sample design in our simulations is well beyond the scope of this study. Nevertheless, we specify a sampling procedure that replicates the pattern of sampling errors

[3]The mean weighted poverty rate equals 12.0 percent, while the mean unweighted poverty rate equals 12.2 percent. The median weighted poverty rate equals the median unweighted poverty rate of 11.7 percent. Both standard deviations equal 3.9 percent, and both interquartile ranges equal 4.7 percentage points. The range of the weighted estimates is 20.2 percentage points, while the range of the unweighted estimates is 20.4 percentage points.

[4]For specifying a population to use in the simulations, it does not appear that there is any loss from ignoring the weights. However, the weighted poverty rates are a limited standard by which to judge the unweighted poverty rates. The weighted poverty rates are fairly unreliable sample estimates and may not accurately reflect the rates that prevailed in 1989.

Table of Contents

TABLE A.3 *(continued)*

Division/State	Assumed Population Size	Assumed Poverty Rate (Percent)	Target Standard Error[a] (Percent)	Sample Size for Simulations[b]
East South Central				
Kentucky	1,630	15.8	2.0	287
Tennessee	1,812	16.6	1.9	318
Alabama	1,860	17.2	2.0	297
Mississippi	2,063	22.1	2.1	339
West South Central				
Arkansas	2,000	16.8	2.0	306
Louisiana	1,525	23.4	2.3	270
Oklahoma	1,774	13.1	1.8	307
Texas	8,772	18.8	1.0	1,327
Mountain				
Montana	2,035	14.4	1.8	321
Idaho	2,093	11.7	1.6	338
Wyoming	1,417	9.7	1.8	231
Colorado	1,690	11.7	1.7	283
New Mexico	2,459	17.6	1.9	337
Arizona	1,886	14.2	1.8	316
Utah	1,949	7.3	1.3	330
Nevada	1,614	9.8	1.6	273
Pacific				
Washington	1,835	9.4	1.5	304
Oregon	1,609	11.1	1.7	269
California	14,413	13.3	0.7	2,231
Alaska	2,122	11.3	1.7	295
Hawaii	1,515	12.7	1.8	272

[a]The target standard error is the standard error for the weighted poverty rate for 1989, estimated from the March 1990 CPS.

[b]The sample size is calculated so that a simple random sample of the indicated size will imply a standard error for an estimated poverty rate generally equal or very close to the target standard error. The expression for calculating sample sizes is given in the text.

Table of Contents

$$(3) \quad n_i = \frac{T_i \, [s_i^2 + p_i \, (1 - p_i)]}{T_i \, s_i^2 + p_i \, (1 - p_i)}.$$

We calculate s_i using the generalized variance function (GVF) estimated by the Census Bureau. The form of the GVF is:

$$(4) \quad s_i = \sqrt{\frac{f_i^2 b}{T_{w,i}} \, p_{w,i} \, (1 - p_{w,i})} \, ,$$

where $p_{w,i}$ is the weighted CPS poverty rate estimate (expressed as a proportion) for state i, $T_{w,i}$ is the base for this estimated poverty rate (the weighted state population), and f_i and b are GVF parameters estimated by the Census Bureau, with values provided in CPS technical documentation. Wolter (1985) discusses the specification, estimation, and limitations of GVFs.

According to Equations 1 and 3, if the sample estimate for a particular iteration is equal to the true poverty rate for state i, the standard error for that sample estimate is exactly equal to s_i. Moreover, it is easy to show that the standard error will be very close to s_i unless the sample poverty rate estimate differs from the true value by many percentage points.[5] Thus, the pattern of standard errors for sample estimates implied by our simple sample design is similar to the pattern of standard errors implied by the complex CPS sample design.

In Table A.3, we display the values of T, p, and s for each state and the implied sample sizes, that is, the values for n calculated according to Equation 3.[6] State sample sizes in our simulations range from about 220 to over 2,200.

---

[5]If the sample poverty rate estimate differs from the true value by 10 percentage points, the standard error will generally differ from s_i by less than 1 percentage point.

[6]The values for p and s in Table A.3 must be divided by 100 before applying Equation 3. Differences between the displayed values of n and the values of n calculated from the displayed values of T, p, and s may arise due to rounding.

Table of Contents

Step 2b: Draw, Without Replacement, a Simple Random Sample of Size n_i for State i, i = 1, 2, ..., 51

The 51 state samples constitute a single national sample. That sample is a stratified simple random sample. Individuals in the population are stratified by state, and independent simple random samples of individuals are drawn in each state.[7]

Step 2c: Draw 1,000 Samples

We repeat Step 2b 1,000 times, drawing 1,000 independent samples. Each of the 1,000 repetitions of our simulation procedure beginning with the drawing of a sample (Step 2b) and ending with the calculation of sample and shrinkage estimates (Step 3) is an "iteration."

STEP 3: CALCULATE SAMPLE AND SHRINKAGE ESTIMATES

Not counting the pooled sample estimates, we calculate 1,000 sets of sample and shrinkage estimates of state poverty rates, one set of 51 sample estimates and one set of 51 shrinkage estimates per iteration. To derive shrinkage estimates, we use an Empirical Bayes shrinkage estimator that combines sample and regression estimates. This estimator was used by Schirm, Swearingen, and Hendricks (1992) to derive state estimates of poverty, FSP eligibility, and FSP participation. Prior to calculating shrinkage estimates, we must calculate sample estimates and their standard errors and specify the regression model to be used.

---

[7]To draw a sample of n_i individuals for state i, we use the SAS function RANUNI. We draw a random number uniformly distributed on the interval (0,1). Multiplying the random number by T_i and adding 1 to the product, we obtain a random number uniformly distributed on the interval $(1, T_i+1)$. Then, we truncate the transformed random number to obtain a discrete random number uniformly distributed over the integers $\{1, 2, ..., T_i\}$. We repeat these steps until n_i unique random numbers are obtained. For example, to select a sample for Maine, we generate 285 unique random numbers distributed over the integers from 1 to 1,603. Those numbers index the individuals selected for that sample. Thus, if 13 is drawn, the 13th individual is included in the sample for Maine.

Table of Contents

Step 3a: Calculate the Sample Estimates

For state i, the sample estimate of the proportion poor is the number of individuals in the sample who are poor divided by the sample size, n_i. Expressed as a percentage, the poverty rate is the proportion poor multiplied by 100. We calculate standard errors for the sample estimates using Equation (1), which gives the standard error for the estimated proportion poor. Multiplying the standard error for the estimated proportion poor by 100 gives the standard error for the estimated poverty rate.

Step 3b: Select the Best-Fitting Regression Model

As described in Chapter I, our regression model regresses the 51 sample estimates of state poverty rates on symptomatic indicators. The symptomatic indicators measure state characteristics that are likely to be associated with interstate differences in poverty rates. Although we do not need to calculate regression estimates prior to calculating shrinkage estimates, we do need to specify the symptomatic indicators that are included in the "best-fitting" regression model in a particular iteration.[8,9] From a set of potential symptomatic indicators, we will include those for which the

---

[8]As shown in Step 3c, we calculate shrinkage estimates using an expression that incorporates the estimation of the best-fitting regression model.

[9]Although the purpose of this study is to compare the accuracy of sample and shrinkage estimates, we report in Chapter III selected results pertaining to the relative accuracy of regression estimates. The expression for our regression estimator is:

$$Y_r = X(X'DX)^{-1}X'DY_s,$$

where X, D, and Y_s are defined under Step 3c. Our regression estimator weights observations by the inverses of the standard errors for the sample estimates of state poverty rates. The variance-covariance matrix of our regression estimator is:

$$V_r = \left[\frac{(Y_s - Y_r)'D(Y_s - Y_r)}{51 - K}\right] X(X'DX)^{-1}X',$$

(continued...)

Table of Contents

model obtained is parsimonious and provides a good fit. Thus, we will not include symptomatic indicators that improve the fit only marginally. We seek a model that accounts for much of the interstate variation in poverty rates with a small number of symptomatic indicators.

We allow for up to five symptomatic indicators: (1) the proportion of the state population receiving SSI, (2) state per capita total personal income, (3) the state crime rate, (4) a dummy variable equal to one for the New England states, and (5) a dummy variable equal to one if at least 1 percent of the state's total personal income is derived from the oil and gas extraction industry.[10,11] Our model-fitting procedure selects the model that maximizes:

$$(5)\quad \bar{R}^2 = 1 - \left[\frac{51 - 1}{51 - k - 1}\right] (1 - R^2) ,$$

where k is the number of symptomatic indicators in the regression model (ranging from one to five), and R^2 is the usual coefficient of multiple determination. Whereas the addition of a symptomatic indicator always increases R^2, $\bar{R}^2$ will decrease if the improvement in fit, as measured by R^2, is small.[12] We repeat our model-fitting procedure for each iteration.

[9](...continued)
where K is the number of variables (symptomatic indicators plus an intercept) in the regression model.

[10]Schirm, Swearingen, and Hendricks (1992) examined these and other symptomatic indicators.

[11]Data on the number of persons receiving SSI are from Table 9.B1, "Number of Persons Receiving Federally Administered Payments and Total Amount of Payments, by Reason for Eligibility," in U.S. Department of Health and Human Services (1990, p. 299). Data on total personal income and total personal income derived from the oil and gas extraction industry are from Table 1, "Total and Per Capita Personal Income by State and Region, 1985-90," and Table 3, "Personal Income by Major Source and Earnings by Industry, 1988-90," in U.S. Department of Commerce (1991c, pp. 30 and 32-41). Data on crime rates (number of violent and property crimes per 100,000 persons) are from Table 294 "Crime Rates by State, 1985 to 1989, and by Type, 1989," in U.S. Department of Commerce (1991b, p. 177). For constructing the first two symptomatic indicators, state resident population totals are from Table 26, "Resident Population--States and Puerto Rico: 1960 to 1990," in U.S. Department of Commerce (1991b, pp. 20-21).

[12] $\bar{R}^2$ adjusts R^2 for the degrees of freedom used to fit the model.

Step 3c: Calculate the Shrinkage Estimates

We use an Empirical Bayes shrinkage estimator. This estimator was used by Ericksen and Kadane (1985, 1987) to estimate population undercounts in the 1980 census for 66 areas covering the entire U.S. and by Schirm, Swearingen, and Hendricks (1992) to estimate state poverty rates, FSP eligibility counts, and FSP participation rates. It was originally developed by DuMouchel and Harris (1983) based on the pioneering work of Lindley and Smith (1972).

The expression for our shrinkage estimator is:

$$(6) \quad Y_c = \left[D + \frac{1}{u^2} P \right]^{-1} D Y_s,$$

where Y_c is a (51×1) vector of shrinkage estimates, and Y_s is a (51×1) vector of sample estimates. D is a (51×51) diagonal matrix with diagonal element (i,i) equal to one divided by the variance (standard error squared) of the sample estimate for state i. $P = I - X(X'X)^{-1}X'$ is a (51×51) matrix, where I is a (51×51) identity matrix (all diagonal elements equal one, and all other elements equal zero) and X is a $(51 \times K)$ matrix containing data for each state on a set of $k = K - 1$ symptomatic indicators.[13] u^2 is a scalar measuring the interstate variability in the sample estimates of poverty rates not explained by the symptomatic indicators. Thus, u^2 reflects the lack of fit of the regression model. We estimate u^2 by maximizing the following likelihood function with respect to u:

$$(7) \quad L = |W|^{1/2} \, |X'WX|^{-1/2} \exp\left[-\frac{1}{2} Y_s' S Y_s \right],$$

where $W = (D^{-1} + u^2 I)^{-1}$ and $S = W - WX(X'WX)^{-1}X'W$. $|W|^{1/2}$ is the square root of the determinant of W. The variance-covariance matrix of our shrinkage estimator is:

[13]The other column of X consists of all ones and allows for an intercept in the regression model.

Table of Contents

$$(8) \quad V_c = \left[D + \frac{1}{u^2} P \right]^{-1} .$$

Standard errors of the 51 state shrinkage estimates are given by the square roots of the diagonal elements of V_c, a (51×51) matrix.[14]

STEP 4: COMPARE THE RELATIVE ACCURACY OF SAMPLE AND SHRINKAGE ESTIMATES

We compare the relative accuracy of the sample and shrinkage estimates according to a wide variety of accuracy criteria, including root mean squared errors (RMSEs) and mean absolute errors (MAEs). An RMSE is the square root of the average squared deviation between the estimates and the true values. An MAE is the average absolute deviation between the estimates and the true values. For all assessments of accuracy, the true poverty rates are the poverty rates in the population specified in Step 1.

As we discuss in Chapter III, we can calculate a RMSE (or MAE) for a given state by aggregating errors across iterations, or we can calculate a RMSE (or MAE) for a given iteration by aggregating errors across states. The RMSE for state i is:

$$(9) \quad \text{RMSE}_i = \sqrt{\frac{\sum_{j=1}^{1,000} (\hat{p}_{ij} - p_i)^2}{1,000}} ,$$

[14]The "final answer" from a Bayesian analysis is a distribution for the true values that we are trying to estimate. The distribution is conditional on the observed data (sample estimates and symptomatic indicators). Our shrinkage estimator, Y_c, is the mean of such a distribution, and V_c is the variance-covariance matrix of the distribution. Given certain assumptions, which were made by DuMouchel and Harris (1983) and Ericksen and Kadane (1985) and which we also make, the distribution is normal. The distribution characterizes the uncertainty that remains after the observed data are taken into account.

Table of Contents

where $\hat{p}_{ij}$ is the estimated poverty rate, p_i is the true poverty rate, and iterations are indexed by j.[15] The MAE for state i is:

$$(10)\quad \text{MAE}_i = \frac{\sum_{j=1}^{1,000} |\hat{p}_{ij} - p_i|}{1,000} .$$

Because states are different sizes, aggregating errors across states for a given iteration raises the issue of how to weight the state errors. If state errors are equally weighted, a one percentage point error in the estimate for a small state will make the same contribution to the RMSE (or MAE) as a one percentage point error in the estimate for a large state, even though the error for the small state may have virtually no impact on, for example, the estimate of the national poverty rate. Alternatively, we could differentially weight state errors, giving greater weight to the errors for large states. Thus, the RMSE for iteration j is:

$$(11)\quad \text{RMSE}_j = \sqrt{\sum_{i=1}^{51} w_i(\hat{p}_{ij} - p_i)^2} ,$$

where w_i is the weight for state i. The MAE for iteration j is:

$$(12)\quad \text{MAE}_j = \sum_{i=1}^{51} w_i |\hat{p}_{ij} - p_i| .$$

We consider three weighting schemes: (1) weighting states equally, (2) weighting states by population shares, and (3) weighting states by poverty shares. When state errors are weighted

[15]When we aggregate errors across iterations for a given state, we can decompose the mean squared error--the RMSE squared--for a state into the sum of the bias squared and the standard deviation squared. The bias of an estimator is the mean error. Although its relevance to an evaluation of accuracy is limited, we do report state-specific biases for each estimator in Appendix B. For a given estimator, the bias for state i is:

$$\text{bias}_i = \frac{\sum_{j=1}^{1,000} (\hat{p}_{ij} - p_i)}{1,000} = \frac{\sum_{j=1}^{1,000} \hat{p}_{ij}}{1,000} - p_i .$$

Table of Contents

equally, w_i = 1/51 for all states, and we get a conventional mean of squared or absolute errors. The population share weights and the poverty share weights are displayed in Table A.4. The population share weight for state i is obtained by dividing the true state i population by the true U.S. population. In other words, it is the share of all individuals in the population specified in Step 1 living in state i. The poverty share weight for state i is obtained by dividing the true state i poverty count by the true U.S. poverty count. In other words, it is the share of all poor individuals in the population specified in Step 1 living in state i. With the population share weights, errors for states with more people are weighted more heavily, while with the poverty share weights, errors for states with more poor people are weighted more heavily. States with more people also tend to have more poor people, so the population share and poverty share weights are closely associated.

In addition to calculating RMSEs and MAEs by aggregating estimation errors across iterations or across states, we calculate these measures of error by aggregating across all iterations and all states. Our expressions for the RMSE and MAE are:

$$(13)\quad \text{RMSE} = \sqrt{\sum_{i=1}^{51} w_i \sum_{j=1}^{1,000} \frac{(\hat{p}_{ij} - p_i)^2}{1,000}}$$

and

$$(14)\quad \text{MAE} = \sum_{i=1}^{51} w_i \sum_{j=1}^{1,000} \frac{|\hat{p}_{ij} - p_i|}{1,000} .$$

POOLED SAMPLE ESTIMATION

To obtain pooled sample estimates, we must add to the first three steps of our simulation procedure. In Step 1, we must define "populations" from which to draw samples.[16] To simulate the most often used procedure of pooling three consecutive annual samples, we use the nonoverlapping

[16]Because many individuals enter the U.S. population (through birth and immigration) and many individuals exit the U.S. population (through death and emigration) during any three-year period, the concept of a population for pooled sample estimation is not well-defined. State-to-state migration and changing family composition present further conceptual difficulties.

Table of Contents

TABLE A.4

WEIGHTS USED TO CALCULATE ROOT MEAN SQUARED ERRORS AND MEAN ABSOLUTE ERRORS

Division/State	Population Share Weight[a]	Poverty Share Weight[b]
New England		
Maine	0.011	0.010
New Hampshire	0.009	0.008
Vermont	0.009	0.008
Massachusetts	0.040	0.036
Rhode Island	0.009	0.009
Connecticut	0.009	0.009
Middle Atlantic		
New York	0.080	0.074
New Jersey	0.042	0.039
Pennsylvania	0.043	0.041
East North Central		
Ohio	0.042	0.041
Indiana	0.010	0.011
Illinois	0.040	0.041
Michigan	0.041	0.040
Wisconsin	0.013	0.013
West North Central		
Minnesota	0.011	0.010
Iowa	0.012	0.012
Missouri	0.011	0.011
North Dakota	0.013	0.013
South Dakota	0.014	0.014
Nebraska	0.013	0.012
Kansas	0.012	0.012
South Atlantic		
Delaware	0.009	0.009
Maryland	0.010	0.010
District of Columbia	0.008	0.009
Virginia	0.015	0.015
West Virginia	0.011	0.012
North Carolina	0.041	0.039
South Carolina	0.013	0.014
Georgia	0.012	0.011
Florida	0.047	0.050

Table of Contents

TABLE A.4 *(continued)*

Division/State	Population Share Weight[a]	Poverty Share Weight[b]
East South Central		
Kentucky	0.011	0.010
Tennessee	0.012	0.011
Alabama	0.011	0.012
Mississippi	0.013	0.013
West South Central		
Arkansas	0.012	0.013
Louisiana	0.010	0.010
Oklahoma	0.012	0.011
Texas	0.050	0.056
Mountain		
Montana	0.012	0.013
Idaho	0.013	0.013
Wyoming	0.009	0.009
Colorado	0.011	0.011
New Mexico	0.013	0.016
Arizona	0.012	0.012
Utah	0.013	0.012
Nevada	0.010	0.010
Pacific		
Washington	0.012	0.012
Oregon	0.010	0.010
California	0.085	0.091
Alaska	0.011	0.013
Hawaii	0.010	0.010

[a]The population share weight is obtained by dividing the "true" state population by the "true" U.S. population.

[b]The poverty share weight is obtained by dividing the "true" state poverty count by the "true" U.S. poverty count.

Table of Contents

TABLE A.4

WEIGHTS USED TO CALCULATE ROOT MEAN SQUARED ERRORS AND MEAN ABSOLUTE ERRORS

Division/State	Population Share Weight[a]	Poverty Share Weight[b]
New England		
Maine	0.011	0.010
New Hampshire	0.009	0.008
Vermont	0.009	0.008
Massachusetts	0.040	0.036
Rhode Island	0.009	0.009
Connecticut	0.009	0.009
Middle Atlantic		
New York	0.080	0.074
New Jersey	0.042	0.039
Pennsylvania	0.043	0.041
East North Central		
Ohio	0.042	0.041
Indiana	0.010	0.011
Illinois	0.040	0.041
Michigan	0.041	0.040
Wisconsin	0.013	0.013
West North Central		
Minnesota	0.011	0.010
Iowa	0.012	0.012
Missouri	0.011	0.011
North Dakota	0.013	0.013
South Dakota	0.014	0.014
Nebraska	0.013	0.012
Kansas	0.012	0.012
South Atlantic		
Delaware	0.009	0.009
Maryland	0.010	0.010
District of Columbia	0.008	0.009
Virginia	0.015	0.015
West Virginia	0.011	0.012
North Carolina	0.041	0.039
South Carolina	0.013	0.014
Georgia	0.012	0.011
Florida	0.047	0.050

Table of Contents

for the pooled sample estimate by multiplying the standard error for the single sample estimate by $\sqrt{0.5}$.[19]

[19]According to Equation (1), if we ignore the finite population correction (fpc), $1 - (n_i/T_i)$, doubling the sample size multiplies the standard error by

$$\sqrt{\frac{1}{2} - \frac{1}{2(2n_i - 1)}} \, ,$$

which very nearly equals $\sqrt{1/2}$ for the values of n_i in our simulations. Because the population from which we draw the pooled sample is not well-defined, we do not adjust the fpc.

observations from the March 1989 and March 1991 CPS samples, ignoring the weights on observations and excluding unrelated individuals under age 15.[17] From these nonoverlapping observations, we draw stratified simple random samples for each iteration. In Step 2, we draw a sample of $n_i/2$ individuals from the March 1989 CPS observations and a sample of $n_i/2$ individuals from the March 1991 CPS observations for state i.[18] These n_i additional individuals are pooled with the n_i individuals selected from the March 1990 CPS. Thus, the pooled sample estimate is based on twice as many observations as the single sample estimate. Population and sample sizes for each of the three years pooled are displayed in Table A.5. Poverty rates for each of the three years are displayed in Table A.6 with the weighted average poverty rates obtained when the populations are pooled. In Step 3, the pooled sample estimate of the proportion poor is the number of individuals in the pooled sample who are poor divided by the sample size, $2n_i$. We estimate the standard error

[17]To determine the poverty status of individuals in the population based on the March 1989 CPS, we averaged OMB poverty income guidelines for the first and last six months of 1988 to obtain calendar year 1988 guidelines. (The annual income data collected in the March 1989 CPS pertain to 1988.) For residents of Alaska, the poverty guideline is $7,035 for a one-person family. Each additional family member increases the guideline by $2,415. For residents of Hawaii, the poverty guideline is $6,480 for a one-person family, and each additional family member increases the guideline by $2,220. For residents of the other states and the District of Columbia, the poverty guideline is $5,635 for a one-person family, and each additional family member increases the guideline by $1,930. To determine the poverty status of individuals in the population based on the March 1991 CPS, we averaged OMB poverty income guidelines for the first and last six months of 1990 to obtain calendar year 1990 guidelines. (The annual income data collected in the March 1991 CPS pertain to 1990.) For residents of Alaska, the poverty guideline is $7,660 for a one-person family. Each additional family member increases the guideline by $2,615. For residents of Hawaii, the poverty guideline is $7,050 for a one-person family, and each additional family member increases the guideline by $2,405. For residents of the other states and the District of Columbia, the poverty guideline is $6,130 for a one-person family, and each additional family member increases the guideline by $2,090.

[18]If n_i is odd, we draw $(n_i + 1)/2$ individuals from one CPS and $(n_i - 1)/2$ individuals from the other CPS. Which sample size was rounded up was determined randomly.

TABLE A.5 *(continued)*

Division/State	Assumed Population Sizes[a] Year 1	Year 2	Year 3	Sample Sizes for Simulations[b] Year 1	Year 2	Year 3
East South Central						
Kentucky	920	1,630	859	144	287	143
Tennessee	906	1,812	936	159	318	159
Alabama	953	1,860	967	148	297	149
Mississippi	1,054	2,063	1,040	170	339	169
West South Central						
Arkansas	977	2,000	1,042	153	306	153
Louisiana	879	1,525	716	135	270	135
Oklahoma	883	1,774	819	154	307	153
Texas	4,297	8,772	4,443	664	1,327	663
Mountain						
Montana	1,015	2,035	978	160	321	161
Idaho	940	2,093	1,112	169	338	169
Wyoming	633	1,417	726	116	231	115
Colorado	834	1,690	924	142	283	141
New Mexico	1,123	2,459	1,162	168	337	169
Arizona	980	1,886	833	158	316	158
Utah	908	1,949	988	165	330	165
Nevada	852	1,614	886	136	273	137
Pacific						
Washington	850	1,835	978	152	304	152
Oregon	770	1,609	743	134	269	135
California	3,972	14,413	7,448	1,116	2,231	1,115
Alaska	1,167	2,122	1,059	148	295	147
Hawaii	732	1,515	651	136	272	136

[a]The Year 2 assumed population size is the assumed population size used for simulating single sample estimation. It is the unweighted number of persons in the March 1990 CPS. The Year 1 and Year 3 assumed population sizes are the unweighted numbers of persons in the March 1989 CPS and the March 1991 CPS living in households that were not in the March 1990 CPS.

[b]The Year 2 sample size is the sample size used for simulating single sample estimation. The Year 1 and Year 3 sample sizes were set equal to one-half the Year 2 sample size. One of the two (Year 1 or Year 3) sample sizes was rounded up, and the other was rounded down if the Year 2 sample size is odd. Which sample size was rounded up was determined at random.

Table of Contents

TABLE A.5

POPULATION AND SAMPLE SIZES FOR SIMULATING POOLED SAMPLE ESTIMATION

	Assumed Population Sizes[a]			Sample Sizes for Simulations[b]		
Division/State	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
New England						
Maine	728	1,603	787	142	285	143
New Hampshire	678	1,340	490	116	232	116
Vermont	638	1,259	582	117	234	117
Massachusetts	2,882	5,745	2,795	526	1,052	526
Rhode Island	676	1,343	572	120	239	119
Connecticut	641	1,365	687	116	233	117
Middle Atlantic						
New York	3,398	11,687	5,882	1,050	2,101	1,051
New Jersey	3,018	6,226	3,107	552	1,105	553
Pennsylvania	3,184	6,488	3,321	570	1,141	571
East North Central						
Ohio	3,269	6,518	3,404	550	1,099	549
Indiana	907	1,769	808	135	270	135
Illinois	3,259	6,486	3,188	528	1,056	528
Michigan	2,965	6,332	3,108	541	1,082	541
Wisconsin	1,025	2,065	1,059	166	331	165
West North Central						
Minnesota	883	1,577	782	138	277	139
Iowa	915	1,884	958	162	325	163
Missouri	885	1,722	787	148	297	149
North Dakota	1,081	1,996	1,037	174	349	175
South Dakota	1,075	2,161	954	180	359	179
Nebraska	955	1,945	1,070	166	331	165
Kansas	839	1,896	984	155	310	155
South Atlantic						
Delaware	693	1,447	658	125	250	125
Maryland	762	1,532	665	128	257	129
District of Columbia	667	1,390	542	108	217	109
Virginia	1,097	2,326	1,120	192	383	191
West Virginia	904	1,841	972	148	297	149
North Carolina	2,847	6,105	2,960	537	1,074	537
South Carolina	1,011	2,175	947	178	355	177
Georgia	881	1,773	867	154	308	154
Florida	3,627	7,869	3,981	618	1,235	617

TABLE A.6 *(continued)*

Division/State	Year 1	Year 2	Year 3	Weighted Average
East South Central				
Kentucky	17.4	15.8	18.4	16.8
Tennessee	17.7	16.6	15.8	16.7
Alabama	17.1	17.2	21.4	18.2
Mississippi	29.1	22.1	25.3	24.6
West South Central				
Arkansas	16.4	16.8	19.6	17.4
Louisiana	22.6	23.4	25.3	23.7
Oklahoma	17.0	13.1	13.8	14.2
Texas	19.4	18.8	17.0	18.5
Mountain				
Montana	15.0	14.4	15.1	14.7
Idaho	11.0	11.7	12.0	11.6
Wyoming	8.4	9.7	14.9	10.7
Colorado	12.0	11.7	14.1	12.4
New Mexico	19.9	17.6	23.5	19.6
Arizona	13.8	14.2	13.8	14.0
Utah	9.6	7.3	7.3	7.9
Nevada	6.6	9.8	10.9	9.3
Pacific				
Washington	7.6	9.4	8.3	8.7
Oregon	11.6	11.1	8.1	10.5
California	13.8	13.3	14.7	13.8
Alaska	13.7	11.3	11.7	12.0
Hawaii	15.2	12.7	11.4	13.0

NOTE: The Year 2 poverty rates are the true poverty rates in the simulations. The Year 1 and Year 3 poverty rates are the poverty rates in the populations from which samples are drawn for pooling with the sample for Year 2. The weighted average poverty rate is obtained by giving weights of 1/4, 1/2, and 1/4 to the poverty rates for Years 1, 2, and 3, respectively.

TABLE A.6

POVERTY RATES IN THE POOLED POPULATION

Division/State	Year 1	Year 2	Year 3	Weighted Average
New England				
Maine	15.7	9.6	11.7	11.6
New Hampshire	5.6	7.3	7.1	6.8
Vermont	7.8	7.1	10.7	8.2
Massachusetts	8.4	8.6	9.3	8.7
Rhode Island	10.2	6.6	6.6	7.5
Connecticut	3.9	3.0	8.9	4.7
Middle Atlantic				
New York	14.3	13.9	14.9	14.2
New Jersey	6.4	7.9	9.1	7.8
Pennsylvania	9.2	9.8	10.9	9.9
East North Central				
Ohio	13.6	9.8	9.6	10.7
Indiana	8.3	12.6	14.6	12.0
Illinois	13.0	11.9	13.7	12.6
Michigan	10.8	12.6	14.1	12.5
Wisconsin	7.9	8.0	9.0	8.2
West North Central				
Minnesota	14.5	11.2	15.2	13.0
Iowa	9.0	10.1	9.9	9.8
Missouri	10.1	11.9	12.1	11.5
North Dakota	12.6	12.1	12.7	12.4
South Dakota	13.7	12.0	14.0	12.9
Nebraska	8.1	11.5	8.1	9.8
Kansas	9.7	10.0	9.5	9.8
South Atlantic				
Delaware	4.0	9.2	7.6	7.5
Maryland	11.9	8.6	7.2	9.1
District of Columbia	15.4	18.2	19.7	17.9
Virginia	7.4	10.9	9.6	9.7
West Virginia	17.1	14.8	19.9	16.6
North Carolina	12.5	11.4	13.6	12.2
South Carolina	10.1	15.7	13.7	13.8
Georgia	13.1	14.9	17.8	15.2
Florida	14.2	11.9	14.0	13.0

Table of Contents

APPENDIX B

ADDITIONAL TABLES OF SIMULATION RESULTS

Table of Contents

In this appendix, we present additional tables of simulation results. In Table B.1, we display for each state and both sample estimators the percentage of iterations for which the shrinkage estimate is more accurate than the sample estimate. Such findings should be interpreted cautiously. We report these and the other state-specific results in this appendix only to show how the effects of shrinkage might vary from state to state, not to forecast the effect of shrinkage for any particular state. In Table B.2, we display RMSEs and MAEs for states. Ratios of shrinkage RMSEs and MAEs to sample RMSEs and MAEs are presented in Table B.3. The percentage changes in RMSEs and MAEs due to shrinkage that we reported in Chapter III can be calculated from these ratios. A ratio of 0.80 indicates a 20 percent reduction in the RMSE or MAE. When there are many estimates of a particular quantity, for example, 1,000 estimates of a state's poverty rate, we can decompose the mean squared error (MSE) of the estimates into the sum of the bias of the estimates squared plus the standard deviation of the estimates squared.[1] In Table B.4, we display state-specific biases and standard deviations for the single sample, pooled sample, and shrinkage estimators. Frequency distributions of absolute biases are shown in Table B.5. According to Table B.5, the median bias of the single sample estimator is roughly 0, as expected. The median bias of the shrinkage estimator is just under 0.3 percentage points, and the median bias of the pooled sample estimator is just over 0.6 percentage points. While the biases in the shrinkage estimator are attributable to regression toward the mean, the source of the biases in the pooled sample estimator can be found in Table A.6 in Appendix A. The pooled sample estimator is an unbiased estimator of the weighted average of the poverty rates for the three years that we are pooling. However, because poverty rates generally change--often substantially--from year to year, that weighted average is different from the poverty rate for the middle year, the year for which we seek an estimate. As shown in Table A.6 and as confirmed by Table B.5, many of the differences are large. In the last table in this appendix, Table B.6, we display confidence interval coverage rates for states.

[1]The MSE is the RMSE squared, and the bias is the average error. An expression for calculating bias is given in Appendix A.

Table of Contents

TABLE B.1 *(continued)*

Division/State	Sample Single	Sample Pooled
East South Central		
Kentucky	86.2	67.1
Tennessee	66.6	53.6
Alabama	88.4	68.6
Mississippi	57.2	76.1
West South Central		
Arkansas	85.9	65.9
Louisiana	34.0	27.4
Oklahoma	35.4	39.8
Texas	39.6	35.4
Mountain		
Montana	29.5	26.6
Idaho	64.0	50.2
Wyoming	41.7	44.0
Colorado	57.3	50.9
New Mexico	79.9	81.0
Arizona	43.5	36.9
Utah	34.1	27.5
Nevada	88.4	64.2
Pacific		
Washington	55.5	52.2
Oregon	91.7	67.9
California	43.0	55.1
Alaska	69.3	57.9
Hawaii	45.3	37.4
Median	57.2	57.1

NOTE: The shrinkage estimate is more accurate than the sample estimate if the shrinkage estimate is closer to the true poverty rate in absolute value.

Table of Contents

Table B.1

PERCENTAGE OF ITERATIONS FOR WHICH SHRINKAGE ESTIMATE IS MORE ACCURATE THAN SAMPLE ESTIMATE, BY STATE

	Sample	
Division/State	Single	Pooled
New England		
Maine	76.4	79.2
New Hampshire	28.8	23.1
Vermont	54.6	60.7
Massachusetts	57.4	41.4
Rhode Island	47.1	51.7
Connecticut	54.1	86.2
Middle Atlantic		
New York	49.8	51.4
New Jersey	58.6	37.5
Pennsylvania	61.3	42.4
East North Central		
Ohio	54.2	75.0
Indiana	36.8	35.3
Illinois	55.2	63.7
Michigan	58.8	41.9
Wisconsin	30.4	22.1
West North Central		
Minnesota	45.4	66.3
Iowa	89.9	57.1
Missouri	95.7	62.9
North Dakota	45.6	45.3
South Dakota	63.7	56.6
Nebraska	47.5	82.2
Kansas	88.6	58.6
South Atlantic		
Delaware	84.9	78.8
Maryland	74.8	58.2
District of Columbia	31.6	22.9
Virginia	84.1	74.3
West Virginia	84.4	78.4
North Carolina	36.2	57.6
South Carolina	90.8	82.8
Georgia	87.4	61.9
Florida	47.2	82.9

Table of Contents

TABLE B.2 *(continued)*

Division/State	Root Mean Squared Error			Mean Absolute Error		
	Sample			Sample		
	Single	Pooled	Shrinkage	Single	Pooled	Shrinkage
East South Central						
Kentucky	1.972	1.742	1.083	1.577	1.420	0.859
Tennessee	1.832	1.278	1.129	1.466	1.016	0.907
Alabama	2.004	1.826	1.082	1.588	1.464	0.842
Mississippi	1.985	2.931	1.497	1.572	2.577	1.207
West South Central						
Arkansas	1.980	1.606	1.082	1.605	1.286	0.861
Louisiana	2.408	1.737	2.339	1.917	1.382	2.062
Oklahoma	1.829	1.721	1.925	1.470	1.396	1.634
Texas	0.974	0.744	1.008	0.792	0.590	0.816
Mountain						
Montana	1.800	1.328	1.969	1.455	1.061	1.756
Idaho	1.611	1.127	1.098	1.297	0.897	0.874
Wyoming	1.804	1.635	1.753	1.431	1.305	1.459
Colorado	1.712	1.402	1.290	1.368	1.108	1.049
New Mexico	1.893	2.544	1.121	1.487	2.211	0.870
Arizona	1.760	1.217	1.537	1.409	0.970	1.272
Utah	1.341	1.080	1.593	1.050	0.858	1.324
Nevada	1.611	1.251	0.966	1.276	1.002	0.748
Pacific						
Washington	1.546	1.300	1.183	1.233	1.059	0.960
Oregon	1.779	1.348	0.986	1.425	1.104	0.769
California	0.663	0.674	0.642	0.533	0.557	0.522
Alaska	1.648	1.423	1.226	1.300	1.152	0.954
Hawaii	1.820	1.296	1.454	1.456	1.026	1.221

Table of Contents

Table B.2

ROOT MEAN SQUARED ERRORS AND MEAN ABSOLUTE ERRORS, BY STATE

	Root Mean Squared Error			Mean Absolute Error		
	Sample			Sample		
Division/State	Single	Pooled	Shrinkage	Single	Pooled	Shrinkage
New England						
Maine	1.585	2.365	1.065	1.270	2.078	0.844
New Hampshire	1.518	1.156	1.743	1.226	0.926	1.517
Vermont	1.496	1.577	1.248	1.186	1.291	1.013
Massachusetts	0.772	0.577	0.696	0.615	0.464	0.555
Rhode Island	1.521	1.430	1.372	1.209	1.145	1.121
Connecticut	1.038	1.899	0.964	0.810	1.695	0.771
Middle Atlantic						
New York	0.721	0.621	0.660	0.571	0.501	0.525
New Jersey	0.734	0.535	0.677	0.592	0.426	0.547
Pennsylvania	0.772	0.584	0.688	0.618	0.473	0.552
East North Central						
Ohio	0.836	1.088	0.739	0.674	0.941	0.595
Indiana	1.822	1.403	1.634	1.438	1.130	1.414
Illinois	0.913	1.018	0.771	0.730	0.853	0.614
Michigan	0.892	0.620	0.745	0.722	0.500	0.603
Wisconsin	1.406	0.980	1.723	1.110	0.782	1.476
West North Central						
Minnesota	1.687	2.253	1.349	1.351	1.927	1.117
Iowa	1.553	1.143	0.962	1.254	0.917	0.771
Missouri	1.740	1.254	0.968	1.373	1.003	0.753
North Dakota	1.514	1.153	1.245	1.194	0.920	1.023
South Dakota	1.595	1.449	1.121	1.284	1.161	0.908
Nebraska	1.559	1.963	1.243	1.252	1.721	1.029
Kansas	1.562	1.143	0.947	1.242	0.911	0.742
South Atlantic						
Delaware	1.635	1.979	0.970	1.300	1.722	0.765
Maryland	1.593	1.235	1.044	1.282	0.995	0.834
District of Columbia	2.393	1.640	2.460	1.916	1.299	2.179
Virginia	1.475	1.519	0.965	1.186	1.285	0.769
West Virginia	1.930	2.312	1.103	1.541	1.961	0.880
North Carolina	0.889	1.040	0.952	0.699	0.880	0.775
South Carolina	1.804	2.237	1.017	1.458	1.947	0.804
Georgia	1.830	1.359	1.021	1.479	1.075	0.815
Florida	0.825	1.314	0.805	0.655	1.182	0.643

Table of Contents

TABLE B.3 *(continued)*

State	Ratio of Shrinkage RMSE to Sample RMSE		Ratio of Shrinkage MAE to Sample MAE	
	Single	Pooled	Single	Pooled
East South Central				
Kentucky	0.549	0.622	0.544	0.605
Tennessee	0.616	0.884	0.619	0.893
Alabama	0.540	0.592	0.530	0.575
Mississippi	0.754	0.511	0.768	0.469
West South Central				
Arkansas	0.546	0.673	0.536	0.669
Louisiana	0.971	1.347	1.076	1.492
Oklahoma	1.053	1.119	1.111	1.170
Texas	1.035	1.355	1.031	1.384
Mountain				
Montana	1.094	1.482	1.207	1.655
Idaho	0.682	0.974	0.674	0.974
Wyoming	0.972	1.073	1.019	1.118
Colorado	0.753	0.920	0.767	0.947
New Mexico	0.592	0.441	0.585	0.393
Arizona	0.873	1.263	0.903	1.311
Utah	1.188	1.475	1.261	1.543
Nevada	0.600	0.772	0.586	0.747
Pacific				
Washington	0.765	0.910	0.779	0.906
Oregon	0.554	0.732	0.540	0.697
California	0.969	0.953	0.979	0.937
Alaska	0.744	0.862	0.734	0.828
Hawaii	0.799	1.122	0.838	1.190
Median	0.800	0.862	0.835	0.841

Table of Contents

Table B.3

ROOT MEAN SQUARED ERROR AND MEAN ABSOLUTE ERROR RATIOS, BY STATE

	Ratio of Shrinkage RMSE to Sample RMSE		Ratio of Shrinkage MAE to Sample MAE	
State	Single	Pooled	Single	Pooled
New England				
Maine	0.672	0.450	0.665	0.406
New Hampshire	1.149	1.507	1.237	1.637
Vermont	0.834	0.791	0.854	0.785
Massachusetts	0.902	1.206	0.903	1.198
Rhode Island	0.902	0.960	0.927	0.979
Connecticut	0.929	0.508	0.952	0.455
Middle Atlantic				
New York	0.916	1.063	0.920	1.048
New Jersey	0.923	1.266	0.923	1.284
Pennsylvania	0.891	1.177	0.894	1.168
East North Central				
Ohio	0.884	0.679	0.883	0.633
Indiana	0.897	1.164	0.983	1.251
Illinois	0.844	0.757	0.841	0.720
Michigan	0.836	1.203	0.835	1.205
Wisconsin	1.226	1.759	1.330	1.887
West North Central				
Minnesota	0.800	0.599	0.827	0.580
Iowa	0.620	0.842	0.615	0.841
Missouri	0.556	0.772	0.548	0.750
North Dakota	0.823	1.080	0.857	1.112
South Dakota	0.702	0.773	0.707	0.782
Nebraska	0.797	0.633	0.822	0.598
Kansas	0.606	0.828	0.598	0.815
South Atlantic				
Delaware	0.593	0.490	0.589	0.444
Maryland	0.655	0.845	0.651	0.838
District of Columbia	1.028	1.500	1.137	1.678
Virginia	0.654	0.635	0.648	0.599
West Virginia	0.571	0.477	0.571	0.449
North Carolina	1.071	0.915	1.108	0.880
South Carolina	0.564	0.455	0.551	0.413
Georgia	0.558	0.751	0.551	0.758
Florida	0.975	0.613	0.982	0.544

TABLE B.4 *(continued)*

Division/State	Bias[a] Sample Single	Bias[a] Sample Pooled	Bias[a] Shrinkage	Standard Deviation[b] Sample Single	Standard Deviation[b] Sample Pooled	Standard Deviation[b] Shrinkage
East South Central						
Kentucky	-0.078	1.058	-0.076	1.971	1.385	1.081
Tennessee	0.049	0.089	-0.605	1.833	1.275	0.954
Alabama	0.013	1.038	-0.011	2.005	1.503	1.082
Mississippi	0.012	2.543	-0.830	1.986	1.458	1.246
West South Central						
Arkansas	0.076	0.633	0.204	1.979	1.477	1.063
Louisiana	0.010	0.302	-2.004	2.409	1.712	1.207
Oklahoma	0.112	1.149	1.444	1.826	1.282	1.274
Texas	0.025	-0.268	-0.562	0.974	0.694	0.837
Mountain						
Montana	0.002	0.377	-1.720	1.801	1.274	0.958
Idaho	0.015	-0.088	-0.553	1.612	1.124	0.949
Wyoming	0.015	0.935	1.148	1.805	1.341	1.326
Colorado	-0.111	0.637	0.591	1.710	1.249	1.147
New Mexico	0.054	2.117	-0.262	1.894	1.411	1.091
Arizona	0.131	-0.135	-1.122	1.756	1.210	1.050
Utah	-0.086	0.495	1.048	1.339	0.961	1.201
Nevada	-0.016	-0.537	-0.154	1.612	1.130	0.954
Pacific						
Washington	0.010	-0.754	0.583	1.547	1.060	1.030
Oregon	-0.023	-0.615	-0.099	1.780	1.200	0.982
California	-0.007	0.485	0.194	0.663	0.468	0.613
Alaska	-0.041	0.703	-0.414	1.648	1.238	1.155
Hawaii	0.028	0.242	-1.112	1.820	1.274	0.937

[a]A bias is calculated as the difference between the average estimated poverty rate across 1,000 iterations and the "true" poverty rate.

[b]A standard deviation is the standard deviation of the 1,000 poverty rate estimates.

Table of Contents

TABLE B.4

BIASES AND STANDARD DEVIATIONS OF SIMULATED ESTIMATES, BY STATE

Division/State	Bias[a] Sample Single	Bias[a] Sample Pooled	Bias[a] Shrinkage	Standard Deviation[b] Sample Single	Standard Deviation[b] Sample Pooled	Standard Deviation[b] Shrinkage
New England						
Maine	0.054	2.026	-0.278	1.585	1.221	1.029
New Hampshire	-0.008	-0.483	-1.477	1.518	1.051	0.926
Vermont	0.067	1.072	0.573	1.496	1.157	1.109
Massachusetts	0.000	0.147	-0.140	0.772	0.559	0.682
Rhode Island	-0.018	0.877	0.667	1.522	1.130	1.200
Connecticut	0.018	1.683	0.226	1.039	0.880	0.938
Middle Atlantic						
New York	0.022	0.374	-0.144	0.721	0.496	0.644
New Jersey	-0.053	-0.092	0.077	0.732	0.527	0.673
Pennsylvania	0.011	0.134	0.141	0.772	0.569	0.674
East North Central						
Ohio	0.016	0.909	0.195	0.837	0.599	0.713
Indiana	-0.022	-0.576	-1.345	1.823	1.280	0.928
Illinois	0.003	0.771	-0.237	0.914	0.664	0.734
Michigan	-0.017	-0.084	-0.224	0.892	0.614	0.711
Wisconsin	-0.062	0.173	1.271	1.405	0.965	1.165
West North Central						
Minnesota	-0.003	1.850	-0.990	1.688	1.287	0.918
Iowa	-0.023	-0.385	-0.001	1.554	1.077	0.963
Missouri	-0.001	-0.418	-0.050	1.741	1.182	0.967
North Dakota	-0.029	0.276	-0.815	1.514	1.120	0.942
South Dakota	-0.031	0.878	-0.544	1.596	1.154	0.980
Nebraska	-0.024	-1.672	-0.857	1.560	1.029	0.901
Kansas	0.040	-0.240	-0.132	1.562	1.118	0.938
South Atlantic						
Delaware	0.004	-1.661	0.125	1.636	1.076	0.962
Maryland	-0.034	0.453	0.296	1.593	1.149	1.001
District of Columbia	0.050	-0.337	-2.126	2.394	1.606	1.239
Virginia	-0.020	-1.157	-0.210	1.476	0.986	0.942
West Virginia	0.057	1.858	-0.023	1.930	1.378	1.103
North Carolina	0.036	0.822	0.538	0.888	0.637	0.785
South Carolina	-0.052	-1.876	-0.105	1.804	1.219	1.012
Georgia	-0.082	0.192	0.025	1.829	1.347	1.021
Florida	0.056	1.171	0.293	0.824	0.596	0.750

Table of Contents

TABLE B.5

FREQUENCY DISTRIBUTIONS OF ABSOLUTE BIASES, BY ESTIMATOR

	Number of States		
	Sample		
Absolute Bias (Percentage Points)[a]	Single	Pooled	Shrinkage
0.0 - 0.1	48	4	8
0.1 - 0.2	3	5	9
0.2 - 0.3	0	4	9
0.3 - 0.5	0	10	1
0.5 - 0.7	0	5	9
0.7 - 1.0	0	8	4
1.0 - 1.5	0	6	8
1.5 - 2.0	0	6	1
2.0 - 2.5	0	2	2
> 2.5	0	1	0

[a]The common boundary of two intervals falls in the lower interval. Thus, "0.1" falls in the "0.0 - 0.1" interval, not the "0.1 - 0.2" interval.

TABLE B.6 *(continued)*

Division/State	Percentage of All 95-Percent Confidence Intervals Including the True Value		
	Sample		
	Single	Pooled	Shrinkage
East South Central			
Kentucky	94.1	88.0	98.4
Tennessee	95.6	95.8	97.0
Alabama	94.6	88.0	98.8
Mississippi	94.8	59.6	94.9
West South Central			
Arkansas	95.6	91.8	98.9
Louisiana	94.9	93.2	80.3
Oklahoma	95.1	85.3	81.4
Texas	95.7	94.1	91.2
Mountain			
Montana	94.4	94.2	75.8
Idaho	93.7	94.8	95.6
Wyoming	94.3	88.9	88.8
Colorado	93.9	91.8	96.1
New Mexico	94.3	66.4	97.8
Arizona	96.4	95.8	89.5
Utah	93.0	92.5	83.2
Nevada	93.4	91.6	97.5
Pacific			
Washington	95.0	88.2	96.0
Oregon	94.9	92.4	97.8
California	94.7	83.2	94.9
Alaska	96.1	91.8	96.2
Hawaii	94.3	94.8	91.7